v.17, 1961 (Locks #1)
405
W891

WORD

JOURNAL OF THE LINGUISTIC CIRCLE OF NEW YORK

EDITORS

ROBERT AUSTERLITZ WILLIAM DIVER ANDRÉ MARTINET

ARTICLES

The Chronology of Certain Sound Changes in Common Slavic as Evidenced by Loans from Vulgar Latin ..... *Charles E. Bidwell* 105

Grammatical Neutralization in Slavic Expressive Forms ..... *Edward Stankiewicz* 128

Le ton et la grammaire dans le Saramaccan ..... *Jan Voorhoeve* 146

Baltic Nominal Systems ..... *William R. Schmalstieg* 164

Grammatical Analysis of the English Comparative Construction ..... *R. B. Lees* 171

Phonological Rules of a Subdialect of English ..... *Joseph R. Applegate* 186

Communication in Poetry ..... *Iván Fónagy* 194

DISCUSSION

Some Methodological Remarks on Generative Grammar ..... *Noam Chomsky* 219

VOLUME 17, NUMBER 2 AUGUST 1961

LINGUISTIC CIRCLE OF NEW YORK

Cercle Linguistique de New York

HONORARY PRESIDENT: **Judah A. Joffe,** Ecole libre des Hautes Etudes

PRESIDENT: **James Macris,** Rutgers–The State University

VICE-PRESIDENTS: **Marshall D. Berger,** The City College of New York
André Martinet, Sorbonne and Ecole pratique des Hautes Etudes
Uriel Weinreich, Columbia University

SECRETARY and TREASURER: **John P. Hughes,** St. Peter's College and Columbia University

EXECUTIVE COMMITTEE: The preceding and

Arthur Abramson, Haskins Laboratories
Robert Austerlitz, Columbia University
William Diver, Columbia University
Elliott Van Kirk Dobbie, Columbia University
Eugene Dorfman, University of Washington
Robert A. Fowkes, New York University and Columbia University
Paul L. Garvin, Ramo-Wooldridge Laboratories
E. Adelaide Hahn, Hunter College
Fredrick H. Jungemann, Columbia University
William G. Moulton, Princeton University
Leo Pap, New York State College, New Paltz
George C. Pappageotes, Columbia University
Laurence Urdang, New York University
Francis J. Whittield, University of California

ADVISORY COMMITTEE FOR FOREIGN COUNTRIES:

Alfredo Barrera Vásquez, Mexico
Emile Benveniste, France
Giacomo Devoto, Italy
†**J. R. Firth,** England
Henri Frei, Switzerland
J. Duchesne-Guillemin, Belgium
B. Havránek, Czechoslovakia
Louis Hjelmslev, Denmark
Bertil Malmberg, Sweden
J. Mattoso Camara, Brazil
†**Paul Menzerath,** Germany
Anton Reichling, The Netherlands
James St. Clair-Sobell, Canada
Alf Sommerfelt, Norway
Antonio Tovar, Spain

WORD, the journal of the Linguistic Circle of New York, is published three times a year in April, August, and December.

Subscription to WORD is $7.50 per year, this including membership in the Circle and postage anywhere in the world. For full-time students of linguistics, the subscription price is $5.00. Dues for membership in the Circle without subscription to the journal are $1.50 a year. Single numbers of the journal, $3.00 each.

Applications for membership, subscriptions to WORD, and orders for back numbers should be sent to the Secretary-Treasurer, **John P. Hughes,** St. Peter's College, 2641 Hudson Blvd., Jersey City 6, New Jersey.

Manuscripts of articles and monographs should be sent either to André Martinet (3, place de la Gare, Sceaux [Seine], France) or to Robert Austerlitz or William Diver (Box 14, Philosophy Hall, Columbia University, New York 27, N.Y., U.S.A.). **Books for review or listing** and **manuscripts of reviews** should be sent to William Diver at the above address. **Exchange journals** should be sent to André Martinet at the address given above.

The editors of the journal are jointly responsible for its contents. Authors of articles are allowed 100, and authors of reviews 25, reprints gratis; they may get additional copies at cost.

CHARLES E. BIDWELL

The Chronology of Certain Sound Changes in Common Slavic as Evidenced by Loans from Vulgar Latin

1.0. This paper will attempt an evaluation of loanwords (especially place-names) from the Vulgar Latin of the Balkan peninsula and adjoining areas with a view to dating certain sound changes of the late Common Slavic period. While the evidence adduced here and the conclusions based upon it relate, strictly speaking, only to the speech of those Slavs who moved southward into the Balkans (the linguistic forerunners of the present-day South Slavs), it may be assumed that the same developments were proceeding with a roughly parallel chronology in the other Slavic dialects.[1]

1.1. Comparative linguistics permits the reconstruction of the following phonemic system for the late Indo-European dialect from which Slavic evolved.[1]

[1] Loans between Slavic and the Finnic and Baltic languages, while offering much more fragmentary information, appear to bear this assumption out. See particularly Valentin Kiparsky's "The earliest contacts of the Russians with the Finns and Balts" in *Oxford Slavonic Papers III* and the same author's "Chronologie des relations slavo-baltes et slavo-finnoises" in *Revue des Etudes Slaves* XXIV (1948), and K. Būga's "Die Vorgeschichte der Aistischen Stämme im Lichte der Ortsnamenforschung" in the *Streitberg-Festgabe* (Leipzig, 1924).

My interpretation of the Indo-European phonemic system is based, particularly insofar as the velar series are concerned, mainly on W. Lehmann, *Proto-Indo-European phonology*, Austin, 1952. As it represents a late stage of the variety of Indo-European from which Slavic developed, laryngeals and reduced vowels are absent, having merged with other phonemes by this time. The much disputed question as to whether an exclusively Balto-Slavic proto-dialect ever existed or whether there was merely a period of parallel Baltic and Slavic development characterized by geographic contiguity and consequent close contacts between Baltic and Slavic speakers is not touched upon in

264873

Obstruents:	p	t	$\widehat{k}$	k				
	b	d	$\widehat{g}$	g	s			
	bh	dh	$\widehat{g}$h	gh				
Resonants:	i	u	m	n	r	l	Vowels:	a e o
Tonal accents:	´	ˆ	Syllabic length:	¯				

Resonants could function either as consonants or, like vowels, as bearers of syllabicity. Syllabic length could occur with the vowels and resonants. Tonal accent was acute or circumflex if it fell on a long syllabic or diphthong. If the accent fell on a short syllabic, the acute-circumflex contrast was neutralized.[2] Before voiced consonants *s* had a voiced allophone [z].

1.2. At an undetermined period *s* after *i u r* or *k* and before a vowel had an allophone [ş], which allophone later evolved to [x], having attained phonemic status through some development as yet obscure, possibly through the entry of a new *s* (from $\widehat{k}$) into the system, or the adoption of loanwords with *x* in position other than after *i u r k* (cf. *xlěbŭ* 'bread' < Gmc. *xlaib-*).[3] The aspiration component of the voiced aspirate stops was lost, so that *bh dh* $\widehat{g}$*h gh* fell together with *b d* $\widehat{g}$ *g*. The $\widehat{k}$ and $\widehat{g}$ became respectively *s* and *z*, merging with the respective allophones of earlier *s* and establishing *z* as a separate phoneme. The *o* and *a* coalesced into one phoneme, which we shall write *å*. Various of the foregoing changes were shared or paralleled by other branches of the Indo-European family. The resultant phonemic system we may term proto-Slavic.

vowels:	i	u					vowel length:	¯
	e	å					tonal accents:	´ ˆ
consonants:	j	w	m	n	l	r		
	p	t	s	k	x			
	b	d	z	g				

this paper. Certainly the late Indo-European dialect sketched in **1.1** is also the ancestor of Baltic. Almost all the developments outlined in **1.2** are also shared by Baltic. Kuryłowicz has placed the consonantal *j*-changes mentioned in **1.3** in the period of Balto-Slavic common development (see his *L'apophonie en Indo-Européen*, Wrocław, 1956, pages 235–236).

[2] There is reason to suppose that in the variety of Indo-European underlying Baltic and Slavic the length contrast was neutralized in diphthongs (vowel plus resonant) but that the former length contrast left its reflex in that long vowel plus resonant became a diphthong with acute accent, whereas short vowel plus resonant became a diphthong with circumflex accent. The phonemic status of accent remained the same into and through the Common Slavic period we are here considering, though the position of the accent underwent a number of complex and, in their details, not clearly reconstructed shifts. Since the position and nature of the Common Slavic accent is not pertinent to the argument of this paper, no further mention will be made of it.

[3] For an exposition of the phonemic development of Indo-European *s* and $\widehat{k}$, $\widehat{g}$, $\widehat{g}$*h* in Slavic, see André Martinet, *Economie des changements phonétiques* (Berne, 1955), pp. 240–242.

The earlier syllabic resonants were resolved into sequences of *i* or *u* plus the respective consonantal resonant.

1.3. There followed a number of changes characteristic of Slavic alone. It is these changes which concern us here. These include changes conditioned by *j* preceding a vowel and by *j* following a consonant. Through the vocalic *j*-changes back vowels following *j* were replaced by front vowels. Thus, *jå* > *je* and *ju* > *ji*. Through the consonantal *j*-changes a host of new phonemes arose: *kj* > *č*, *gj* > *ž* (through a stage [ǯ]), *xj* and *sj* > *š* *zj* > *ž*, *tj* > *ť*, *dj* > *ď*, *lj* > *l'*, *nj* > *ń*, and *rj* > *ŕ*.

Analogous to the consonantal j-changes are the affrications, the so-called "palatalizations," which affected the velars. The so-called first palatalization gave the same reflexes as the consonantal *j*-changes, namely, *k g x* before proto-Slavic front vowels became *č ǯ š*. The so-called second and third palatalizations show identical reflexes, namely, *k* > *c*, *g* > *ʒ*, *x* > *ś* (*ś* falls together with *s* in South and East Slavic and with *š* in West Slavic). The second palatalization occurred when a velar was followed by a front vowel resulting from the monophthongization of a proto-Slavic diphthong. The third palatalization occurred when a velar was preceded by a front vowel under certain complicated (and not thoroughly defined) conditions.[4]

[4] I offer the following hypothesis regarding the phonological development of the reflexes of the velars before *j* and front vowels. In their initial stage, *č* and *ǯ*, resulting from consonantal *j*-changes and the so-called first palatalization, were allophones of *k* and *g* ([k g] before back vowels, something like [č' ǯ'] before front vowels). The passage of *ě* after *š ž č ǯ* to *a* would have been sufficient to establish *č* and *ǯ* as phonemes, but this probably occurred later than the passage of *dj* to *ě* (or *i* in certain positions), which established *č* and *ǯ* as phonemes, for now [k] and [g] (probably fronted varieties of these consonants) again appeared before the front vowels *ě* and *i*. Presumably the new allophones of *k* and *g* before these front vowels were at first highly fronted and then later assibilated, becoming [ć ǯ] eventually. A corresponding allophone of *x* eventually became [ś] and then coalesced with *s* in East and South Slavic dialects and with *š* in West Slavic dialects. In various parts of the Slavic domain, *c* and *ʒ* (from the allophones [k' > ć] and [ǵ > ǯ] attain phonemic status in various ways during or immediately after the period when the Slavs began to migrate. In the West Slavic area they first coalesced with the reflexes of *t'* and *d'*, which also originally appeared only before front vowels, but later, as the result of the backing of certain front vowels or analogical substitution of back vowel for front vowel in certain paradigms came to stand before both front and back vowels. In a large part of the South Slavic area (the dialects from which Slovenian, Serbo-Croatian, and Macedonian developed) the reflexes of *t'* and *d'* took the place of *c* and *ʒ* as allophones of *k* and *g* before front vowels, since the former were phonetically closer to *k* and *g*, being then probably [t' d'] or [k' g'] (cf. the reflexes of *t'* and *d'* in modern Macedonian, namely *k' g'*). The reflexes of *t'* and *d'* continued to function as allophones of *k* and *g* until the coalescence of *y* with *i* in South Slavic and of *ŭ* with *ĭ* in western South Slavic caused *k* and *g* to appear before front vowels again. In this connection, note

The proto-Slavic vowel system underwent a thorough refashioning, whereby quality replaced length as a phonemic distinguishing feature (though length probably remained as a phonetic fact). Thus: *ī* > *i*, *ū* > *y*, *ē*, *ā* > *a*, *i* > *ĭ*, *u* > *ŭ*, *e* > *e*, *â* > *o*. At the same time, the diphthongs are simplified (monophthongized): *ej* > *i*, *âj* > *ě*, *âw* > *u*, *ew* > *ü*, a back vowel plus *m* or *n* followed by a consonant became *ǫ* (a back nasal vowel), and a front vowel plus *m* or *n* followed by a consonant became *ę* (a front nasal vowel). The resulting phonemic system is that of late Common Slavic:

vowels:	i	ū	y	u				tonal accents:	´	^
	ĭ			ŭ						
	ę			ǫ						
	e		o							
	ě		a							
consonants:	p		t				k			
	b		d				g			
			s			š	x			
	v		z			ž		r	ŕ	
			c		t′	č				
			ʒ		d′	ǯ		l	l′	
	m		n		ń					

Note that the earlier semivowel *w* (written *v* at this stage in accord with traditional practice) now functions as a labial spirant. The phoneme *ü* was probably [ju] or [jü] (consonants preceding it underwent *j*-changes).[5]

Trubetzkoy's statement in "Die altkirchenslavishe Vertretung der urslavishen **tj*, **dj*," *Zeitschrift für slavische Philologie* XIII (1936), pp. 88–97: "Im Urserbokroatishen müssen die aus **tj*, **dj* entstandenen palatalen Verschlusslaute ebenfalls als kombinatorische Varianten der Velaren k g gewertet werden. Zu selbständingen Phonemen werden sie erst nach Aufheben der Gegensätze *u-ü*, *y-i*, *a-ä*."

The relative chronology of the so-called second and third palatalizations has been much discussed. Recently F. V. Mareš in his article "Vznik slovanskeho fonologického systému a jeho vývoj do konce období slovanské jazykové jednoty" in *Slavia* XXV (1956), see especially pages 464–468, has renewed the argument that the so-called third palatalization preceded the second. Whatever the merits of Mareš's arguments (some of his reasoning seems quite compelling, but other parts weak or even self-contradictory) regarding the priority of the third palatalization, he is surely right in asserting that the first stage of the processes traditionally termed second and third palatalizations was an allophonic fronting of the velars (Mareš holds, correctly I believe, that when the velars were fronted in the third palatalization, back vowels following them were replaced by the corresponding front vowels, the fronted velars thus remaining allophones not contrasting with the unfronted velars before back vowels), that the results of both palatalizations were at this stage already phonemically and phonetically identical and that the final stage (assibilation or affrication) was therefore undergone simultaneously by the fronted velars resulting from both second and third palatalizations.

[5] For a discussion of the phonemic analysis of Old Church Slavic, which in its oldest stages is almost identical with the stage here designated Common Slavic (the only

To complete the picture, it should be mentioned that word-final consonants fell, intervocalic consonant clusters were simplified, and the sequences vowel (other than *ĭ* or *ŭ*) plus *r* or *l* before consonant were either metathesized (South and West Slavic) or given an anaptyctic vowel between the *r* or *l* and the following consonant, with the result that all syllables now terminate in a vowel, except (in part of the Slavic domain) those terminating in *ĭ* or *ŭ* plus *r* or *l*.[6] Final vowels and diphthongs gave reflexes differing in certain cases from those outlined above, which are valid for non-final position. After *š ž č ǯ, ě* > *a*. After monophthongization, *j* may be regarded as having lost phonemic status, occurring as an allophonic glide before front vowels after another vowel. Alternatively, if *j* is retained as a phoneme, the stock of phonemes may be reduced, *ŕ l' ń t' d' ü* being resolved into *rj lj nj tj dj ju*. A resolution of *š ž č ǯ* into *s z c ʒ* plus *j* would also be theoretically possible. *ǫ* and *ę* might be resolved into *o* and *e* plus *n* or a component of nasality.[7]

1.4. As stated above, we are here concerned with the changes which occurred between the stage I have termed proto-Slavic and the break-up of the Common Slavic language into a number of independent dialects, a process well under way by the late ninth century after the appearance of the first written records of Slavic and the shattering of the continuity of Slavic-settled territory with the intrusion of the Magyars into Pannonia.

On the basis of internal evidence we may attempt to place certain of the changes in relation to others. All investigators agree that the first palatalization preceded the second and third. The relative chronology of the latter two has been the subject of disagreement. Since both the latter

---

difference being that Common Slavic *t'* and *d'* become *št* and *žd* in Old Church Slavic and that an additional phoneme *jǫ* must be added), see Horace G. Lunt, "On Old Church Slavonic phonemes: the Codex Zographensis" in *Word* VIII (1952), 311–328 (see particularly the discussion of the phonemic status of *j* on pages 318–321) and N. S. Trubetzkoy, *Altkirchenslavische Grammatik*, Sitzungsberichte der Österreichischen Akademie der Wissenschaften, Philosophisch-historische Klasse, vol. CCXXVIII:4, particularly the discussion of nasal phonemes on pages 36 and 81–82.

[6] The treatment of the sequence vowel plus liquid plus consonant varies in the various parts of the Slavic domain. The tendency toward open syllables was not carried through completely in all parts of the domain. In this connection, see Roman Jakobson "On Slavic diphthongs ending in a liquid" in *Word*, VIII (1952), 306–310.

[7] Accounts of the evolution from Indo-European to Common Slavic differing somewhat from the foregoing may be found in: N. van Wijk, "K istorii fonologičeskoj sistemy v obščeslavjanskom jazyke" in *Slavia*, XIX, 293–313; F. V. Mareš, "Vznik slovanského fonologického systému a jeho vývoj do konce období slovanské jazykové jednoty" in *Slavia*, XXV, 443–495; and J. M. Kořínek, *Od indoeuropského prajazyka k praslovančine* (Bratislava, 1948).

palatalizations show identical reflexes, it is, in the absence of contrary evidence, at least possible that they were roughly contemporaneous, rather than one preceding the other as generally assumed by previous investigators. With regard to the first and second palatalizations, it would seem that, since they gave different reflexes for the same sequences of phonemes, not only did the first palatalization begin before the second, but its operation must have terminated before the second began.[8] After the first palatalization terminated and as a prerequisite to the second palatalization, monophthongization must have occurred, by which the sequences velar plus *âj* were converted into velar plus *ě* (or under certain conditions, *i*). Thus, proto-Slavic **kâjna* > **kěna*, resulting in Common Slavic *cěna* 'price' after the second palatalization. Metathesis of liquids occurred after the first palatalization at the earliest, since we have examples like Old Church Slavic *žlědǫ* 'I repay' from proto-Sl. **geld-*. The palatalization must have occurred before the metathesis, for the conditioning factor, namely the front vowel immediately following the velar, would not otherwise have been present. Placing the *j*-changes is somewhat more difficult. There is no guarantee that all the changes lumped together under that heading were effective at the same time. In fact, certain of them continued to be operative well into historic times, in various Slavic dialects, while others did not. Thus, in one variety of Serbo-Croatian, *n* plus *je* resulting from Common Slavic *ě* becomes *ńe*, while *s* plus *je* from the same source does not become **še: nežĭnŭ* > *ńežan* 'gentle, delicate' but *sěverŭ* > *sjever* 'north.' With regard to the consonantal *j*-changes, it would appear that those affecting the velars could not have begun to be operative later than the beginning of the first palatalization, because the reflexes are the same and it is highly unlikely that a velar stop could be affricated by a front vowel and remain unaffected by a following [j]. For reasons of structural similarity, it would appear probable, though by no means certain, that the remaining consonantal *j*-changes began roughly at about the same time as the velar *j*-changes. Regarding the vocalic *j*-changes, it is impossible to determine on internal evidence, whether they began before, after, or at the same time as the consonantal *j*-changes. As to the change of *ě* to *a* after *č*, *ǯ*, *š*, and *ž*, it would appear to have occurred relatively late, perhaps contemporaneously with the liquid metathesis, since examples where the

[8] Some investigators (e.g. van Wijk in the article cited under footnote 5) assume a phoneme *ě*$_2$ < *âj* which caused the second palatalization of preceding velars and then fell together with *ě*$_1$. Such an assumption would permit overlapping of the first and second palatalizations. However, not only is this solution unnecessarily complicated and phonetically unlikely, but it will not take care of cases of second palatalization in loanwords caused by front vowels other than *ě*$_2$, e.g. *cĭrky* (cf. Russian *cerkóv*) 'church' < Greek *kirjakí* or from a Germanic word derived from the Greek etymon.

sequence in question were interrupted by metathesis either do not show the change or show doublets with and without the change. Thus, **čeln-* > *člěnŭ* and *članŭ*.

2.0. When the Slavs began to settle in the Balkans toward the end of the sixth century of our era, they did not find an ethnographic vacuum. The area southward of a line running roughly from the Adriatic at Dyrrachium (Durazzo), leaving Scupi (Skopje), Naissus (Niš), and Remesiana (Bela Palanka) to the North, and then following the crest of the Balkan mountains to the Black Sea near the mouth of the Danube, was Greek speaking (with blocs of aboriginal speakers in Thrace and possibly southern Macedonia.[9]. North of this line they encountered a partially Romanized population with Vulgar Latin speakers along the coast, along the Danube, and probably in most of the rest of the area, with blocs of speakers of aboriginal Indo-European languages (Thracian? Dacian? Illyrian?) scattered in the interior. Our evidence for the presence of aboriginal speakers at the time of the Slavic settlement is the existence of the present-day Albanian language and the fact that certain place-names seem to have been borrowed by Slavic from an aboriginal language rather than from Greek or Latin. Thus, the name of the Bulgarian city Plovdiv is from Thracian Pulpudēva rather than Greek Philippoupolis. The non-Slavic population north of Greece was gradually assimilated or driven into refuge areas. Local Romanic speech survives in the form of Rumanian in Rumania proper and in scattered groups in the Balkans south of the Danube. It survives in heavily Venetianized form in the towns along the western coast of Istria. Its Dalmatian variety survived until the end of the last century on the island of Veglia (Krk), having become extinct earlier in the various Dalmatian cities and islands, persisting longest in Ragusa (Dubrovnik), where it apparently died out in the sixteenth century.

2.1. The Vulgar Latin of the Balkans and neighboring regions was not uniform. There must have been a number of dialect differences. We may distinguish the variety from which Rumanian was derived, which was probably spoken in northern Macedonia and the area on both sides of the present Serbian–Bulgarian border and was similar to, if not identical with,

[9] This delineation of the linguistic situation in the Balkans prior to the Slavic settlement there is based primarily on Konstantin von Jireček's *Die Romanen in den Städten Dalmatiens während des Mittelalters* (published as nos. 48 and 49 in the Denkschriften der kaiserlichen Akademie der Wissenschaften, Vienna, 1902 and 1904) I, 13. I have deviated from this line in including Scupi and at least part of the Haemus in the Latin area (following Skok in "Zum Balkanlatein IV", *Zeitschrift für romanische Philologie*, LIV, 179).

that variety of Vulgar Latin from which the oldest stratum of Vulgar Latin loans in Albanian was derived.[10] We further distinguish the variety from which sprang Dalmato–Romanic (or Old Dalmatian) formerly spoken on the eastern Adriatic coast and islands from the Gulf of Quarnero south to perhaps Durazzo and from which the second stratum of Vulgar Latin loans in Albanian was derived, an Istrian variety spoken on the Istrian peninsula, and proto-Rhaeto-Romanic spoken in Friuli and the Julian region as far south as Trieste and probably in a good part of the Austrian and Slovenian Alps. Our evidence is far too sketchy to posit any further dialect divisions, though other dialects may have existed in such regions of the Balkan interior as were Romanized.[11] Proto-Istro-Romanic and Proto-Rhaeto-Romanic belonged to the western group of Romanic dialects in which intervocalic geminates and some clusters were simplified, intervocalic voiceless stops and *s* voiced, and intervocalic voiced stops became fricatives, while Proto-Dalmato-Romanic and Proto-Rumanian belonged to the eastern group where intervocalic stops are retained as in Classical Latin and intervocalic clusters generally treated more conservatively than in the West.[12] In the greatest part of the Vulgar Latin domain the velar stops *k*

[10] Concerning the location of the area in which contact between proto-Rumanian and proto-Albanian could have taken place, see Henrik Barić, *Lingvističke Studije* (Djela Naučnog Društva NR Bosne i Hercegovine I, Sarajevo, 1954), particularly p. 25 *et seq.* N. van Wijk offers evidence implying that the area along the Serbo-Bulgarian border was long occupied by a Romanic population in his "Taalkundige en historiese gegevens betreffende de oudste betrekkingen tussen Serven en Bulgaren" (*Medeelingen der koninklijke akademie van wetenschappen*, deel 55, series A, Amsterdam, 1923). A discussion of the various theories offered concerning the area or areas in which proto-Rumanian was spoken is contained in M. Friedwagner, "Ueber die Sprache und Heimat der Rumänen in ihrer Frühzeit", *Zeitschrift für romanische Philologie* LIV (1934), 641–675.

[11] E. Kranzmayer in his "Frühromanische Mundarten zwischen Donau und Adria" (*Zeitschrift für Namenforschung* XV) makes some daring hypotheses regarding possible Romanic isoglosses in the Pannonian and inner Balkan area on the basis of the few German and Slavic place-names in this area derived from Latin names.

[12] Skok holds (and brings impressive, though not completely conclusive, evidence to support his thesis, in "Considérations générales sur le plus ancien istro-roman" in *Romanica Helvetica* XX [1943], 472–485) that Istro-Romanic was originally a Romanic dialect of the Eastern type, closely allied to Dalmato-Romanic, which was later overlaid and replaced by a dialect of the Rhaeto-Romanic (West Romanic) type. Skok's evidence consists largely of the Slavic place-names in Istria, of which such as *Capris* > Slovenian *Koper* 'Capo d'Istria', *Petena* ~ *Petina* > SCr. *Pićan* 'Pedena', which are apparently derived from VL etyma with unvoiced intervocalic stops (but note that a number of apparently equally old Istrian place-names in Slavic show voicing of intervocalic stops: *Piquentum* > *Buzet*, 'Pinguente', see 3.3 below for development of this item) and of apparent survivals in present-day Istro-Romanic dialects. R. L. Politzer in "A note on the North Italian voicing of intervocalic stops", *Word* XI (1955), 416–419, adduces evidence from eighth-

and *g* underwent an affrication before the front vowels *e* and *i* analogous to the so-called palatalizations in Slavic. This development occurred only partially in part of the Dalmato-Romanic area and not at all in most of it; *k* or *g* preceding *i* were affricated in the northern extremity of the area, on Veglia, but no further south, as the next island, Arbe (Rab), has a place name *Krklant* 〈 **circinātus*.[13] This difference in treatment is illustrated by the reflex of Latin *cimicem* 'bedbug' 〉 Italian (representative of the treatment common to most of the Romanic area) *cimice* /čimiče/, Vegliote *činko*, Ragusan dialect of Serbo-Croatian (representing a loan from the local Dalmato-Romanic) *kimak*. Rumanian, like most of the Romanic area, shows affrication of velars before front vowels, but there is some doubt as to whether the affrication in Rumanian took place before the arrival of the Slavs in the Balkans.[14] In the entire Vulgar Latin area velar stops before [j] were affricated. [j] also effected changes of other preceding consonants analogous to the Slavic consonantal *j*-changes.

2.2. By the time of the Slavic settlement in the Balkans the Classical Latin three-height vowel system with quantity distinction had given way in most of the Romanic area to a system with four distinctions of height and no phonemic distinction of quantity (though allophonic differences of vowel length determined by position certainly existed).[15] Thus:

Classical Latin			Vulgar Latin			
ī i		u ū	CL ī 〉	i	u	〈 CL ū
ē e	o ō		CL i ē 〉	ę	ǫ	〈 CL u ō
ā a			CL e 〉	e	o	〈 CL o
				a		〈 CL a ā

century documents that voicing of intervocalic stops had not yet taken place at that period in some areas of Northern Italy, particularly in northeastern Italy. If Western Romanic voicing of intervocalic stops spread from a western center (Gaul?) and had not yet covered all northern Italy by the eight century, this would support the thesis that voicing of intervocalic stops had not yet taken place in Istro-Romanic when the Slavs began to settle in Istria (seventh century).

[13] The Vegliote affrication of *k* and *g* before *i* is very probably a late development, as it takes place also before the reflexes of VL *u* (which passed through a stage *ü*). It is surely late if Skok is correct in deriving the place-names on Veglia *Maknel* from **machinile*, *Krknul* from **circinālis*, and *Kakorajne* from **cicerina* see his "Studi toponomastici sull' isola di Veglia" in *Archivio glottologico italiano* XXV, 319–355. See also Carlo Tagliavini, *Le origini delle lingue romane* (Bologna, 1948), p. 265 (cited hereafter as Tagliavini, *Le origini*).

[14] See Tagliavini, *Le origini*, pp. 257–258.

[15] In most of the VL area, vowels under stress were phonetically (nondistinctively) long in open syllables, short in closed syllables. Furthermore, the stressed vowels in antepenultimate syllables were shorter than those in penultimate syllables. It is probable that even short stressed vowels were phonetically slightly longer than unstressed vowels,

Proto-Rumanian however showed an asymetrical vowel system:[16]

CL ī	> i		u	< CL u ū
CL i ĕ	> ę	o	<	CL o ŏ
CL e	> e			
		a	<	CL a ā

In unstressed syllables all varieties of Vulgar Latin apparently had only three vowel height distinctions, the distinctions *ę* versus *e* and *ǫ* versus *o*, if present in a given dialect, being neutralized except under stress.

2.3. Romanic languages which evolved from these varieties of Vulgar Latin and with which the Balkan Slavs were in contact at periods later than that during and immediately following their invasion of the Balkans (and hence not within the scope of this paper) include: Friulan, spoken in north-eastern Italy bordering on the Slovenian speech area; Istriote, formerly spoken in Istria, a last remnant heavily influenced by Venetian surviving in the small city of Rovigno and a few neighboring villages; Old Dalmatian, represented by Vegliote and other extinct dialects of various Dalmatian cities and islands (the last speaker of Vegliote died near the end of the nineteenth century after having served as an informant for the Italian linguist Bartoli); Venetian, an Italian dialect of western Romanic type which replaced local Romanic speech in Trieste and in such parts of Istria and Dalmatia where Romanic speech was not replaced by Slavic; and Rumanian, varieties of which are spoken in Rumania proper and scattered elsewhere in the Balkans.

3.0. For the period we are here treating only the earliest Romanic loans can be considered. These are, in first line, names of places early occupied by Slavs, of geographical features within the Slavic settlement area, as well as the names of places or geographical features, which, though peripheral to areas of Slavic settlement, were known to the early Slavic settlers.[17] In

which were uniformly short. See chapter II of Harald Weinrich's *Phonologische Studien zur romanischen Sprachgeschichte* (Münster, 1958); also C. Tagliavini, *Le origini*, pp. 143 and 145 and C. H. Grandgent, *Introduction to Vulgar Latin* (Boston, 1907), pp. 76–77. I am indebted to André Martinet for the information regarding difference of vowel length between the penult and antepenult in Vulgar Latin.

[16] The Proto-Rumanian vowel system also underlay the Vulgar Latin loans in Albanian. The evidence for Dalmato-Romanic, as reflected in Vegliote, is not entirely clear. A. Colombis in "Elementi veglioti nell'isola di Cherso-Ossero," *Archivum Romanicum* XXI (1937), 243–267, postulates, on the basis of Dalmato-Romanic loans in the Venetian and Serbo-Croatian dialects of Cherso, the same vowel system for Proto-Dalmato-Romanic as for Proto-Rumanian.

[17] A good part of the place-name etymologies adduced as examples in this paper are drawn from the various publications of the late Yugoslav linguist Petar Skok, particu-

citing the modern Slavic derivatives of these loans the ordinary orthography (or, in the case of Bulgarian citations, a transliteration) will be employed, and such features as stress and vowel quantity will not be marked, since these suprasegmental features in the modern languages are irrelevant to the changes here considered. Latin etyma, if proper names, will be cited in their attested orthography. Vowel length will be marked in those other than proper names.

3.1. If we compare modern Slavic items with their Latin etyma, we may reconstruct the following correspondences between the Vulgar Latin original and the phonemic shape it assumed at the stage we designate as Common Slavic. The correspondences listed below are not valid in word-final position, in which position, as in the case of Slavic final vowels inherited from Indo-European, more complicated statements apply.

VL stressed *a* is borrowed as *a* when in the VL penult and as *o* in the VL antepenult. This is an indication that VL stressed antepenult vowels were shorter than stressed penult vowels. *Savus* > *Sava*, *Dravus* > *Drava*, *Pagus* > *Pag*, *Massarum* > *Mosor*, *Apsorus* or *Apsarum* > *Osor*, and *Kotor* compared to It. *Càtaro*.

VL unstressed *a* is borrowed as *o*. *Parentium* > *Poreč*, *Salona* > *Solin*, *ad Musculum* > *Omišalj*, *ad Portulam* > *Oprtalj*.

Evidence is scanty for VL stressed *o*, but would appear to indicate it was borrowed as *ŭ*. *Scodra* (> **skŭdr-* > SCr. *Skadar* 'Scutari') may have had short *o*; Albanian *Shkodër* would seem to indicate this. In position before *r* plus consonant we have *ad Portulam* > **opŭrtŭlĭ* > *Oprtalj*, *Corcyra Nigra* > **kŭrkŭrŭ* > Old SCr. *Krkar* (modern SCr. *Korčula* is from Italian) '*Curzola*.'

Evidence is also insufficient for VL unstressed *o*. Here too, it would appear to have been borrowed as *ŭ*. Classical *Bononia* (VL **bonǫ́nja*) > *Bŭdynŭ* > *Bdin* (medieval name of *Vidin*), *Cormones* (so attested in 610 and 628) = VL *kormǫ́nes* > *Krmin* (Slovenian name for Cormons). And possibly *Corinium* > **kŭrinŭ* > *Karin* (locality in Dalmatia), if Skok is correct in identifying it with ancient Corinium.

larly his "Studije iz ilirske toponomastike" which appeared in the *Glasnik zemaljskoga muzeja u Bosni i Hercegovini*, vols. XXIX, XXXI, and XXXII (Sarajevo, 1917, 1919, and 1920), his "Zum Balkanlatein" in the *Zeitschrift für romanische Philologie*, vols. XLVI, XLVIII, L, and LIV, and his *Slavenstvo i Romanstvo na jadranskim otocima* (two volumes, Zagreb, 1950, hereafter cited as *Slavenstvo i Romanstvo*), as far as the identification of the modern word with its ancient etymon is concerned. On the details of the evolution from the Vulgar Latin etymon to the modern South Slavic form I differ from Skok in a great many cases.

VL stressed *ǫ* and *u* are usually borrowed as *ū* > Common Slavic *y* > South Slavic *i*. *Salona* > **solynŭ* > *Solin* (there are a host of other names in *-ōna*, all of which show modern South Slavic *-in*, e.g. *Scardona* > *Skradin*, *Aenona* > *Nin*, etc.), *Tragurium* > **trogyrŭ* > *Trogir*, *Utinum* > VL **údin-* > **vydĭnŭ* > *Videm* (Slovenian name for Udine).[18]

VL unstressed u is borrowed as *ŭ*. Thus, *ad Musculum* > **omyšŭl'ĭ* > *Omišalj*, *ad Portulam* > **opŭrtŭl'ĭ* > *Oprtalj*.[19]

In some borrowings, VL stressed and unstressed *o*, *ǫ*, and *u* appear in modern South Slavic as *u*. Thus, *Gregorius* > *Grgur*, *Solenta* > *Sulet* (modern SCr. *Šolta* is from Venetian), *Pola* > *Pulj* (modern SCr. *Pula* is again a later borrowing from Venetian), *Doclea* > *Duklja*, *Medeon* > *Medun*. Items showing this reflex apparently belong to a slightly later stratum of borrowings. Of the examples cited above, for instance, the first is a personal name introduced after Christianization. The rest are place-names from the periphery of the area of Slavic settlement, where a non-Slavic population, from whom the name could be borrowed at a later stage, presumably survived longer. (Šolta is an Adriatic island. Pola was predominately Venetian speaking until after 1945. Duklja and Medun are both near the present border between Albania and Montenegro, an area of mixed Slavic, Romanic, and Albanian population in the early middle ages.) In the case of Šolta and Pola (SCr. *Pula*) the older Serbo-Croatian names *Sulet* and *Pulj* have been replaced by patently recent borrowings.

VL stressed *ę* and *i* are usually borrowed as *i*. *Nedinum* > **nĭdinŭ* > *Nadin* (locality in Dalmatia), *Oescus* > VL *ęsku* (*Esco* is attested in the Peutinger tables) > *Iskər* (river in northwestern Bulgaria).

VL unstressed *i* is borrowed as *ĭ*. Thus, *civitātem* > **cĭvĭtatŭ* > *Cavtat*, *Serdica* > *Srědĭcĭ* (medieval name for Sofia).

[18] Latin intervocalic voiceless stops are voiced in the Friulan area, as in all of the Western Romanic domain, the exact time when this occurred being uncertain (see Politzer's note cited under footnote 12). Prothetic *v* is regular before initial back vowels. *m* instead of *n* is a hypercorrection (*m* # > *n* # in Slovenian dialects on the western periphery).

[19] *Ulpiana* > **lypľanŭ* > *Lipljan* (a town in Kosovo-Metohija) is puzzling. Here we have metathesis of the initial liquid, as is expected, but the reflex of the unstressed *u* is that of proto-Slavic long rather than short *u* (the expected development is that of *Ulcinium* > *Lacin*, cf. footnote 29 below). The labial context may have preserved a closer articulation of *u* in *Ulpiana*, thus accounting for the exceptional treatment (personal communication of André Martinet). The treatment *Ulpiana* > *Lipljan* parallels the development **ūrba* > *ryba* 'fish' tentatively proposed by Roman Jakobson in footnote 1 to his article "Slavic Diphthongs Ending in a Liquid" in *Word* VIII (1952), 306. Transmission through or contamination by aboriginal (i.e. non-Romanic) Balkan dialects may possibly account for some of the discrepancies noted.

Though in most cases stressed VL *ę, i* > Common Slavic *i* and stressed VL *ǫ, u* > Common Slavic *y* as stated above, the material examined contains a few examples (all listed below) where the VL phonemes were borrowed as *ĭ* and *ŭ* respectively, having apparently been heard by the Slavs as short and equated with their short *ĭ* and *ŭ*. The following items show these reflexes: *Scupi* > **skŭp+ĭje* > *Skopje* (*ŭ*> *o* in Macedonian), *Curicum* > **kŭrĭkŭ* > *Krk*, *Corcyra Nigra* > **kŭrkŭr-* > Old SCr. *Krkar* (modern SCr. *Korčula*, It. *Curzola*), *Muccurum* > **mŭkŭr-* > *Makar* and *Makarska*, ancient *Vegia*, early medieval *Bigi* (attested in the anonymous geographer of Ravenna) > **bĭg-* > modern SCr. *Bag* or *Karlobag*,[20] *Biston* > **bĭstŭ* > *Bast*, *Cissa* (ancient name for the island *Pag*) > VL *kęssa* (*Chessa* is attested in medieval documents) > **cĭs+ĭska* > *Caska* (name of a locality on *Pag*).[21] It will be noted that in all but three (*Scupi*, *Vegia*, *Corcyra*) of the foregoing items the stressed vowel is either in a closed syllable or in the ante-penult (or both). These are the positions where stressed VL vowels were phonetically shorter than in open penultimate syllables (see footnote 15).

VL stressed *e* is borrowed as *e*. *Senia* > *Senj*, *Céleia* > *Celje*, *cerrus* > SCr. *cer* 'a species of oak, quercus cerrus,' *Neveste* > *Nevest*, VL *venét-* > **bĭnetŭk* > Old SCr. *bnetk-* > modern SCr. *mletk-* 'Venice.'

VL unstressed *e* appears to be borrowed as *ĭ*. Thus, VL *venét-* > **bĭnetŭk-* > SCr. *bnetk-* > *mletk-* 'Venice,' *Céleia* > **celĭj-* > *Celje*, *Nedinum* > **nĭdinŭ* > *Nadin*, *Glemona* > **glĭmynŭ* > **gl̥min* > *Humin* (Slovenian name for Gemona in Friuli; in the Slavic dialects of this area *g* > *h*, vocalic *l* becomes *u*); but in one example, *Neveste* > *Nevest*, we have the reflex *e*.

VL front vowel plus nasal consonant before another consonant results in *ę*. *Parentium* > **porękĭ* > *Poreč* 'Parenzo.'

VL back vowel plus nasal consonant before another consonant results in *ǫ*. *Basssantem* > **bosǫtŭ* > *Bosut*, *juncus* > VL **žǫnku* > **žǫkŭ* > SCr. dialect *žuk* 'reed,' *Aesontius* > **sǫča* > *Soča* 'Isonzo.'

Before *r* or *l* followed by consonant, VL *e* or *ę* and *a* result in *ě* and *a*

[20] Final *g* rather than *z*, which would be expected as a result of the second palatalization has been explained (by Fran Ramovš in "O prvotnih substitucijah za balk.-lat. k, g pred e, i," *Južnoslovenski Filolog* VI, 165) as due to the borrowing of the place name as a locative and the replacement of *z* by *g* in the other cases by analogy with the regular morphophonemic alternation *g*~*ʒ* (later *z*) in Common Slavic. Cf. Old Church Slavic *bogŭ* 'God', *o boʒě* 'about God'.

[21] To these examples of Latin stressed *i* (VL *ę*) > Slavic *ĭ*, all of which are taken from South Slavic, there might be added the following two loans which are widespread in the Slavic domain: *piperem* 'pepper' > **pipĭrĭ* > SCr. *papar*, Czech *pepř*, Polish *pieprz*; *missa* 'mass' > SCr. dialect and Slovenian *maša*, Czech *mše*, Slovak *mša*, Polish *msza*. Unstressed VL *e* > Slavic *ĭ* is also exemplified by *piperem* > **pipĭrĭ*.

respectively in loans after undergoing the Common Slavic metathesis. The reflexes *ě* and *a* are the usual ones for metathesized *e* and *å* in South Slavic.[22] *Arba* > *Rab*, *Scardona* > *Skradin*, *Albona* > *Labin*, *Melita* > VL **mélta* > *Mljet*, *Syrmium* > VL *sẹrm-* (cf. *sermenses* in CIL III 9551) > **srěmŭ* > *Srem*, *Srijem*, *Serdica* > *Srědĭcĭ*.

In a few examples (possibly later borrowings) the sequence vowel plus *r* or *l* does not metathesize and *a* or *e* are replaced by *ŭ* or *ĭ*. Thus, *Tarsatica* > *Trsat*, presumably through a stage **tŭrsatŭ*. Likewise *Bergona* > *Brgin* and *Stelpona* (also attested as *Stalpona*, *Stolpona*) > *Stupin*. In borrowings which occurred later than the period we are considering here foreign *a* or *e* plus *r* is very frequently rendered by South Slavic vocalic *r*. Cf. such patently late SCr. borrowings as *Bernardino* > *Brnardin* and colloquial *mrš* 'get out,' presumably from German *marsch*.

One statement may be made about vowels in word-final position. As is to be expected, Vulgar Latin loans are assimilated to Slavic inflectional types with substitution of Slavic endings or suffixes for Vulgar Latin final vowels. The vast majority of nouns enter the masculine declension with the nominative ending *ŭ* or *ĭ*.

3.2. The treatment of vowels and diphthongs in loans from Vulgar Latin appears to indicate that transition from the proto-Slavic phase with diphthongs and vowel length was in its initial stages at the time of the first Slavic settlements on the Balkans. Thus, in the earliest borrowings Vulgar Latin stressed *u* and *ǫ* result (by way of *y*) in *i* in the South Slavic languages. VL *ǫ*, being long and a higher vowel than proto-Slavic *a*, was heard and

[22] Some investigators (most recently Mareš in the work cited under footnote 4) have held that the changes in Slavic traditionally labeled "metathesis of liquids" involved, not only in East Slavic, but in the whole Slavic domain, the existence at a certain stage of an anaptyctic vowel between the liquid and the following consonant. A certain amount of support for this theory is lent by the names *Srakane* (two small islands west of Cherso; in Italian, *Canidole*), if this name is, as appears possible, derived from *Saracenus* 'Saracen' (Arab corsairs were active in the upper Adriatic at various times in the ninth century and c. 841 a Saracen fleet inflicted a severe defeat upon the Venetians in a sea battle off *Susak*/*Sansego*), and *Mljet* (from Latin *Melita*), if the second vowel in the ancient name had not been already lost in Vulgar Latin by syncope (Constantine Porphyrogenitus in the tenth century has *Méleta*; modern Italian, derived from the Venetian name, is *Méleda*), and by *Aspaláthos* > VL **spelet-* (with the alternation of *e* and *a* found frequently in names of this area; cf. *Spalatum* ~ *Spaletum* in Thomas Archidiaconus, *Delmati* ~ *Dalmati*, *Serdica* ~ *Sardica*, *Méleta* in Constantine Porphyrogenitus ~ *Malata* in the anonymous geographer of Ravenna) > **splětŭ* > *Split* (Spalato) and possibly by *Curicum* > **kŭrĭkŭ* > *Krk* (Veglia). *Lom* < *Almus* (in north-western Bulgaria) does not give the expected South Slavic long vowel reflex and deserves further investigation.

borrowed as proto-Slavic *ū*, as was also VL *u*. If at this time *ǻw* had already been monophthongized to *u* following the unrounding of proto-Slavic *ū* to *y*, VL *u* and *ǫ* would have given *u* in modern South Slavic, as indeed Romanic *o* did in later borrowings (cf. Venetian *roza* > SCr. *ruža*). In an early loan like VL *Salǫna* > SCr. *Solin* the VL *ǫ* was borrowed as *ū* which passed to *y* with the restructuring of the Slavic vowel system (monophthongization and loss of phonemic length) and then to *i* in modern South Slavic.

If VL *ǫ* and *u* were borrowed as *ū* > *y* > *i*, we should expect that *au* or *ou* borrowed at the same period would become *u* in Common Slavic and in later South Slavic.[23] We have one example. *Poetovio* (present-day Slovenian *Ptuj*, German *Pettau*, which, lying in the interior, adjacent to Pannonia,

[23] V. Kiparsky in *Die gemeinslavischen Lehnwörter aus dem Germanischen* (Annales Academiae Scientiarum Fennicae, Series B, Tomus XXXII, Helsinki, 1934) and F. Mareš in the article cited under footnote 4 assume a development *ǻw* [*ō*] > *u*, for which Slavic borrowings from and into Germanic, Baltic, and Finnic appear to argue, e.g. Germanic *plōg-* > OChSl. *plugū*. Kiparsky on p. 285 of the above work says: "Got. *ō* und urgerm. *o* haben stets slav. *u*, niemals *y*, gegeben, weil das heutige slav. *u* noch in urruss. Zeit (etwa um 900) denselben Lautwert wie das germ. *ō* gehabt hat (die Ostseefinnen, die sowohl *ū* als *ō* hatten, wählten zur Wiedergabe der urruss. *u* ihr *ō* . . . und ebenso taten die Letten, wenn sie slav. *duma* durch *duoma* < *dōma* wiedergaben)." The very numerous Slavic borrowings from Balkan Latin on the other hand contradict this hypothesis, for if a phoneme *ō* had been present in the Slavic system when the Slavs entered the Balkans, they would not have borrowed VL *ǫ* as *ū* > *y* > *i* (*Salona* > *Solin*). In general, the borrowings between Slavic and Latin, Germanic, and Finnic corroborate each other's evidence very nicely. On this one point they seemingly do not, and it deserves further investigation. With regard to the loans from Germanic, I venture the following very tentative hypothesis: It is not at all uncommon for long *e* and *o* to become diphthongized, acquiring a high semivocalic offglide in many or all positions (cf. modern English, some dialects of German, and some varieties of colloquial Hungarian). It is quite possible that in those varieties of Germanic, from which Slavic borrowed, long vowels were diphthongal in nature. This would agree nicely with the position of those who regard the long vowels of early Germanic dialects as complex nuclei (see Robert P. Stockwell, "The Phonology of Old English", *Studies in Linguistics* XIII, 1–2, particularly pp. 16–17). Thus, the Slavs might have equated a Germanic long *o* having semivocalic offglide with their own *ǻw*, which at this stage might well have been [oụ]. I am not acquainted with the details of Finnic and Baltic evolution. However, it is evident from the examples quoted by Kiparsky that his "urrussische Zeit" refers to a period after Common Slavic and after monophthongization, since he cites Russian words as giving Finnic *ō* where the *u* in the Russian word is not from proto-Slavic *ǻw*. Thus, Veps *kōma* < *kumū* < Latin *commater*. It is, of course, quite probable that after monophthongization Slavic *u* was relatively open, so that Balts, Finns, Greeks, and others would equate Slavic *u* with their own relatively closer *o* when they borrowed from Slavic, just as the South Slavs after monophthongization (e.g. at the stage when *Duklja* < *Doclea* and *Solun* <*Saloniki* were borrowed) equated Greek and Romanic *o* with their *u*.

was overrun very early by the Slavs),[24] was presumably VL **petǫ̯jo* resulting in Common Slavic **pětuĭ* > *Ptuj*. In other borrowings, VL *au* > modern South Slavic *av* or *ov*. *Laurentius* > *Lovreč*, *aureata* > *ovrata* or *lovrata* (name of a fish; the *l* is the Romanic article). These latter represent borrowings more recent than *Ptuj*, as indicated by the areas of culture involved (Christianity, maritime pursuits).

The following items, which present reflexes of Common Slavic *i* rather than *y* in some or all the Slavic languages, require special comment. Latin *Romam* 'Rome' and *crucem* 'cross' give respectively OChSl. *rimŭ*, Rus. *ŕim*, Pol. *Rzym*, Czech *Řím*, SCr. and Slovenian *Rim* and Pol. *krzyż*, Czech *kříž*, SCr. and Slovenian *križ* (this word, used by Roman Catholics, does not appear in Russian, as the Russians use a different word for 'cross'.) These items were most probably borrowed from a Romanic population somewhere along the Adriatic coast or the Friulan border in the form **rūmu* > **rymŭ* and **krūži* > *kryžĭ* (the latter item was certainly borrowed from a western Romanic variety of speech which voiced intervocalic consonants; cf. Venetian *doge* < Latin *ducem*). In South Slavic, Common Slavic *y* and *i* fell together, giving the reflex *i*. In western South Slavic (i.e., Serbo-Croatian and Slovenian), the reflexes of *ŭ* and *ĭ* also fell together and the allophonic palatalization before front vowels is lost. Apparently, western South Slavic, which borrowed these items, early lost the distinction between *y* and *i*, so that, when these items were transmitted somewhat later to other varieties of Slavic which retained the distinction between *i* and *y*, they were borrowed as *rimŭ* and *križĭ*. This would indicate that coalescence of *i* and *y* (and probably of *ĭ* and *ŭ*) with concomitant loss of allophonic palatalization) had already taken place in at least some varieties of western South Slavic before the borrowing of these words into West Slavic and the appearance of *Rimŭ* in Old Church Slavic texts (hence probably before 1000 A.D.).[25]

[24] Josip Klemenc in *Ptujski grad v kasni antiki* (Academia scientiarum et artium slovenica, opera 4, Ljubljana 1950), pp. 76–77, believes the Slavs settled the area near Poetovio shortly after the Langobards abandoned Pannonia to the Avars (569), probably about 580.

[25] The Freising Monuments, three short passages written in an archaic Slovenian, apparently by German missionary priests around 1000 A.D., indicate that at the time of their composition *ĭ* and *ŭ* had fallen in a weak position and coalesced into one phoneme (probably a schwa-like vowel) in strong position. In two of the three monuments *y* and *i* had also merged (into *i*), while in the other the distinction was preserved after labial consonants with merger taking place in other environments. See Fran Ramovš and Milko Kos, *Brižinski Spomeniki* (Ljubljana, 1937), particularly pp. 7 and 12, and also Fran Ramovš, "Drobnosti iz slovenske gramatike," in *Slavia* I, pp. 27–37. Ramovš believes that the monuments represent copies of originals composed as much as one or even two centuries previously. They would thus represent a stage of the language considerably earlier than 1000 A.D.

At the time of the Slavic invasions of the Balkans the restructuring of the Slavic vowel system must have been in its earliest stages, since the oldest borrowings from Vulgar Latin indicate survival of the diphthong *ȃw* and retention of vowel quantity as phonemically distinctive. Somewhat later *au* begins to be borrowed as *ov* and Romanic *ǫ* and *u*, borrowed as *u*, no longer change to *y*, but are equated with the reflex of proto-Slavic *ȃw*. Romanic unstressed *u* and *i* continue to be borrowed as *ŭ* and *ĭ* and Romanic vowel plus nasal consonant as *ę* and *ǫ* at least until the fall of the weak *jer*s, since even after the restructuring of the Slavic vowel system the respective Romanic and Slavic phonemes were phonetically similar.

As the first palatalization preceded the restructuring of the Slavic vowel system and the second palatalization could not have taken place before the restructuring, we may assume that the second palatalization began to be operative about the time of the Slavic invasion of the Balkans or shortly thereafter. This is borne out by the evidence of the loans, as will be seen in the next section.

3.3. Utilization of Vulgar Latin loan material for the dating of consonant changes is rendered difficult by the fact that Vulgar Latin underwent a number of sound changes similar to the Slavic palatalizations and consonantal *j*-changes. Fortunately for our study, a good part of Balkan Vulgar Latin did not affricate velars before *e* and *i* (at the very least, the Adriatic coast from Veglia to Durazzo). Therefore, any affrication of velars before front vowels in loans made in that area must be attributed to Slavic developments.[26]

The explanation of the change of *y* to *i* in ***rymŭ*, ***kryži* as due to initial borrowing into a Slavic dialect in which *i* and *y* coalesced was, as far as I know, first suggested by Meillet in *Le Slave Commun* (Paris, 1924), p. 102. Vaillant in his *Grammaire comparée des langues slaves* (Lyon, 1958), I, 120, attributes the *y* in ***rymŭ*, ***solynŭ*, etc. to the Rumanian change Latin *o* > *u* before nasal consonant, but this assumption is mistaken, for we have the same borrowing of *-ona* as *-ynŭ* in Romanic areas where *o* definitely did not pass to *u* before nasal consonants (e.g. in Friuli Cormones, VL ***kormǫ́n-*** > Slovenian *Krmin* 'Cormons', *Glemona* > Slovenian *Humin* 'Gemona'), which latter fact indicates that Romanic *o* was borrowed in the South Slavic area as Slavic *ū* > *y* > *i* without an intervening stage of Romanic *u*.

[26] As stated above (see footnote 14), there is considerable doubt whether Rumanian had affricated its velars before front vowels by the time the Slavs reached the Balkans. Latin loans in Albanian show no affrication (Latin *k g* before front vowels > *k′ g′* in Albanian, just as do the reflexes of the Indo-European velars in the same position. See G. Mayer "Die lateinischen Elemente im Albanesischen" in Gröber's *Grundriss der romanischen Philologie* (Strassburg, 1904–06), and "Albanesische Studien III" in volume CXXV of the *Sitzungsberichte der philosophisch-historischen Klasse der kaiserlichen Akademie der Wissenschaften* (Vienna, 1892). Kiparsky's (in the work cited under

As we have concluded that the first palatalization preceded monophthongization and the second palatalization followed that development and as we have further concluded that monophthongization was in process at the time of the Slavic settlement in the Balkans, we may expect to find no cases of the first palatalization in material borrowed by Slavic on Balkan territory, but that any loans from Vulgar Latin with velars followed by front vowel will show the second palatalization reflexes. We have a number of examples with VL *k*. Among place-names we have *Celeia* >

footnote 23) assumption that "romanische Assibilierung" is responsible for the changes *k* > *c* and *g* > *z* in Balkan place names is mistaken (particularly for the Adriatic coastal area for which we have incontravertible evidence that no affrication of velars before front vowels took place).

With due respect for Skok's excellent and voluminous philological and etymological researches it must be stated that his position on a number of details of the evolution of Slavic and Romanic is quite untenable. Skok was taken to task on some of these points by Anton Grad in "Della palatalizzazione di *k* latino intervocalico nel dialetto veneziano," *L'Italia Dialettale* IX (1939), 230–239, and by Fran Ramovš in the article in *Južnoslovenski Filolog* VI cited under footnote 20. Regarding the treatment of velar stops before front vowels in loans from Vulgar Latin, Skok held (see p. 386 of "Zum Balkanlatein II" in *Zeitschrift für romanische Philologie* XLVIII) that velar before front vowel in items borrowed contemporaneously might undergo the first, second, or third Slavic palatalizations or remain unchanged. Ramovš correctly points out that affrication in those items which Skok interprets as examples of first palatalization actually results from other sources and that the items which show unchanged velars before front vowel were borrowed from Dalmato-Romanic after the palatalizations had ceased to be operative in Slavic and after further changes had made it possible for a velar to occur before front vowel. Regarding the items which show the reflexes of the second and third palatalizations Ramovš states that they cannot be the result of the Slavic palatalizations since these had already long since taken place. Rather, these items, according to Ramovš, represent cases where the Slavs in borrowing had substituted their phonemes *c* and *ʒ* for VL positional allophones (before front vowel) [k′] and [g′]. Skok, replying in *Južnoslovenski Filolog* VIII ("Slavenska palatalizacija sa romanističkog gledišta"), accepted this interpretation. I do not. Ramovš offers no evidence whatever that the Slavic palatalizations were operative at a "prehistoric" period long before the Slavic invasions of the Balkans, but simply states it as an accepted and proven fact. More important, Ramovš's explanation is phonetically and phonemically unlikely. As stated in footnote 4 above, by the time the original velars before front vowels had become [ć] and [ʒ] they were no longer allophones of *k* and *g*, but their place as allophones of *k* and *g* before front vowel was taken by the reflexes of proto-Slavic *tj* and *dj*, which were phonetically nearer *k* and *g*, being then probably [t′ d′] or [k′ g′]. Had the sound changes we call second and third palatalization really been long completed as Ramovš maintained, then VL [k′ g′] before front vowel would have been borrowed as the (then) Slavic allophone of *k* and *g* before front vowel, which would have been the reflexes of proto-Slavic *tj* and *dj*, and loans made in this period would show ć and ʒ in modern Serbo-Croatian. This is not the case; early VL loans in SCr. show second-third palatalization reflexes; ergo, second and third palatalization was operative when the linguistic ancestors of present-day Serbo-Croatian speakers settled in the Balkans.

Celje (in Lower Styria), classical *Crexa* or *Crepsa* > VL **kers-* (the VL form is postulated on the basis of an Old Dalmatian name of the island which survives in the modern Italian name *Cherso* first attested in Latin documents of the twelfth century and of the adjective *chersīnus* occurring in Cassiodorus[27]) > *Cres*, *Cissa* > VL *kęssa* > **cĭs+iska* > *Caska*,[28] *civitātem* > **cĭvĭtatŭ* > *Cavtat* (a town in southern Dalmatia, the antique *Epidaurus*; another *civitātem* in an area where *k* was already affricated before front vowels in Vulgar Latin when its name was borrowed is Cividale di Friuli, *Čedad* in Slovenian, *Čabdad* in older Serbo-Croatian), *Cebro* or *Kebron* > *Cibər*, *Cibrica* (town and river respectively in northwest Bulgaria), *cerrus* > SCr. and Bulg. *cer* (designation for a kind of oak, 'quercus cerrus'), *Ulcinium* > **lŭcinĭ* > /lacin/[29] and possibly *cęta* (name of a coin) < *quinta* (a Roman coin).[30] There are fewer examples for *g* before front vowel. *Piquentum* > **pinquentum* > VL **pilgént-* (dissimilation of the first *n* to *l*, loss of *u* and voicing of *k* to *g* in accord with Istro-Romanic development) > **bĭlzętŭ* > *Bl̨zet* > *Buzet* (a town in Istria, in Italian, *Pinguente*),[31] *Genta* > *Zęta* (both *Genta and Zenta* are attested in medieval Latin documents) > *Zeta* (name of a river and district in Montenegro),

[27] Skok in *Slavenstvo i Romanstvo*, p. 35, mentions the occurrence in Cassiodorus without giving any further citation.

[28] Skok's assumption (in *Slavenstvo i Romanstvo*, p. 69) that the present dialect form of the name, *Caska*, is a result of cakavism (cakavism in Serbo-Croatian dialectology signifies the merger of the phonemes *š ž č* with *s z c* in many areas of Dalmatia under the influence of Venetian, which shows a similar coalescence) is mistaken. Rather the original form of the name must have had *c*, and the form *Čaška* which sometimes occurs in the standard language represents a hypercorrection. Cf. also the hypercorrect form *Čres*, sometimes used for *Cres*.

[29] The name of this town, ancient *Ulcinium* (Latin) or *Olkínion* (Greek), now in Albanian speaking territory (though within the political boundaries of Yugoslavia), although it apparently had a Romanic speaking population until late in the middle ages, shows many variants in Serbo-Croatian. *Ulcinj* is the present official form and the following variants are attested: *Ocin*, *Ocinj*, *Ucin*, *Dulcinj*, *Ulćin*, *Lĭcinĭ* (this latter writing is found in medieval Serbian documents from the thirteenth century on and probably represents a pronunciation /lacin/). *Lacin* is probably the native Serbo-Croatian development of this word and the remaining forms show varying degrees of contamination by foreign forms such as Albanian *Ulqin* /ulk'in/ and Italian *Dulcigno*. The native form of the name would then have resulted from metathesis of an original *ŭl-* (cf. *Ulpiana* > *Lipljan* treated above). In any event, important for the present argument is the fact that all the Serbo-Croatian variants of the name show *k* > *c* (excepting only *Ulćin*, which is based on Albanian /ulk'in/ in all probability).

[30] Skok points out that this word has a modern Serbo-Croatian reflex *ceta* 'ancient coin' in the dialects of the Sisak (< *Siscia*) area. Siscia was an important Roman mint and Roman coins are still frequently dug up in this region.

[31] The voicing of the initial consonant in the Slavic reflex is difficult to explain, but the development of the remainder of the word is quite normal.

VL (or Old High German) **gila* > Slovenian *Zilja*, German (Austro-Bavarian) *Gail* (name of a river in Carinthia).[32] As Vulgar Latin had no *x* or *h*, we have no example of the palatalization of *x* in our material. We have a possible example of the third palatalization in a Balkan place-name, namely *Serdica* > *Srědĭcĭ* (medieval name of Sofia), providing that the final *-ĭcĭ* is not the result of substitution of the Slavic suffix *-ĭc-*, quite common in place-names of the Balkans (cf. *Kostolac*, *Aleksinac* etc. in Serbia), in place of the original final sequence of phonemes. Such substitution almost surely took place in *Logatec* (a locality in the Julian Alps corresponding to Roman *Longaticum*), for if palatalization took place in the *-ĭkŭ* of **lǫgatĭkŭ* < *Longaticum*, it is difficult to explain why it did not in *Curicum* > **kŭrĭkŭ* > *Krk*. Better evidence for dating the third palatalization is afforded by the fact that areas of Greece and Austria which had been Slavicized and were then early re-Hellenized or Germanized show place-names of Slavic origin where the third palatalization was not carried through. Thus, **žabinika* > *Sapinicha* (attested 985) > *Sarning* (village and stream on the left bank of the Enns in Upper Austria) and > *Sabinicha* (attested 998) > *Sarming* (village and stream near Perg in Upper Austria). Likewise, **gârdiku* > *Gardíki* and **kamenika* > *Kameníkja* (names of places in the Peloponnesus). The Greek examples are of interest because the *k*'s in them are phonetically palatal (velars in modern Greek have palatal allophones before *j* and front vowels) and may hence indicate the initial stage of the development whose end result we term third palatalization. The German examples indicate that in some areas at least the final result of the palatalization process (affrication) had not been carried through by the time Germanization of the areas in question took place (approximately 650 to 800).[33]

[32] The Gailtal lies north of Friulan speaking territory, thus outside the area where we know VL velars were preserved without affrication before front vowels. However, the German form Gail seems to indicate that this name was borrowed before the affrication of velars had reached this area (as was *Kelsbach* < *Celeusus* in Bavaria). Perhaps the Slavs encountered German speakers (who arrived in the western Austrian Alps shortly before the Slavs began to settle the Balkans) in this area and borrowed the name from them, rather than from Latin. The German form *Gil* (presumably with a long vowel) is attested three times at least in documents from the early twelfth century. Ramovš's derivation (in the article cited under footnote 25) of both the Slavic and German forms of the name from a *Vallis Julia* attested in Paulus Diaconus is unconvincing. The Friulan name *Zeje* seems to be a borrowing or reborrowing from the Slavic. Kranzmayer and Ramovš cite OHG *Gila* without an asterisk, but do not state where it is attested.

[33] The significance of these Austrian place-names in dating the Slavic third palatalization was, to my knowledge, first pointed out by Primus Lessiak and has later been discussed by a number of scholars, including R. Ekblom in his "Germanische und slavische Palatisierung in Ortsnamen des östlichen Alpengebietes," *Språkvetenskapliga Säll-*

From the foregoing we may conclude the third palatalization began at a fairly late date and that its final stage at least was contemporaneous with that of the second palatalization (the reflexes of both are identical). This final stage (affrication) certainly took place after Slavic settlement in the Balkans and had been completed in the dialect(s) on which Old Church Slavic was based before our first documents in that language (end of the ninth century).

3.4. With regard to the metathesis of liquids, we have documentary evidence which dates it fairly precisely, so that the testimony of place-names and other loans merely confirms this evidence. Our documentary evidence (Slavic personal names in Latin documents) shows the first evidence of metathesis in 784 (*Trebel* ⟨ **trěbelĭ* in a document from Salzburg). The name *Karl* (of Charlemagne 771–814) was borrowed by the Slavs with the meaning 'king' and shows metathesis in every Slavic language (SCr. *kralj*, Pol. *król*, Rus. *korólʹ*, etc.). By the time of our first extensive documents in Old Church Slavic metathesis of the liquids was in the main completed. We may infer then that metathesis took place roughly between 750 and 900. Hence, all Vulgar Latin loans taken into Slavic in the period immediately following the settlement in the Balkans show metathesis: *Serdica* ⟩ *Srědĭcĭ*, **kers-* ⟩ *Cres*, *Scardona* ⟩ *Skradin*, *Arba* ⟩ *Rab*, *Albona* ⟩ *Labin*.

3.5. Further, the name **kers-* ⟩ *Cres*, which we treated above under the second palatalization, indicates that this palatalization was operative before metathesis (the change *k* ⟩ *c* could have only taken place before there was an intervening *r* between the *k* and the following vowel).[24] Thus, the

---

skapets i Uppsala Förhandlingar 1949–1951, pp. 37–52. The Greek examples were listed by Max Vasmer in his work *Die Slaven in Griechenland* (Abhandlungen der preussischen Akademie der Wissenschaften, phil.-hist. Klasse nr. 12, Berlin, 1941).

[34] The vowel in *Cres* represents the reflex of Common Slavic *ě* (the local dialect is *ekavski*; neighboring *ikavski* areas call the island *Cris*) as is usual in South Slavic metathesized forms where the initial element of the diphthong had been *e*. With regard to my statement that Cres represents the second and not the first palatalization, it should be noted that the *ča*-dialects of the Quarnero region and neighboring Istria do not show the change *čr* ⟩ *cr* which occurs in the standard language. It is true that the town *Cres* and some other localities on the island have "cakavian" dialects (i.e. all *č*'s are replaced by *c*; see footnote 28), but the name of the island is everywhere *Cres* (or *Cris* in *ikavski* dialects), in all localities, including those with non-cakavian dialects (which ought therefore to show *čr*-, had the original form of the name had that cluster rather than *cr*) on the island and in neighboring regions. The form *Čres* which appears on a few recent maps is nowhere attested in local speech and evidently (like *Čaška* mentioned in footnote 28) represents a hypercorrection made on account of local cakavism (see Skok's

beginning (and probably the termination) of the period during which the second palatalization was operative lies sometime between roughly 600 (beginning of Slavic settlement in the Balkans) and 750 (beginning of metathesis). The second palatalization had certainly ceased operation by the time of our first documents in Slavic, for we have forms like *ķesarĭ* 'emperor, Caesar', *anģelŭ* 'angel', *arxierejĭ* 'high priest'.

4.0. We may now establish the following absolute chronology with regard to some of the late Common Slavic sound changes.

(a) The first palatalization ceased operation before circa 600 A.D. (*terminus ad quem:* Slavic settlement in the Balkans).

(b) Restructuring of the Slavic vowel system, involving monophthongization and replacement of phonemic length by other distinctions, was operative during the earliest period of Slavic settlement in the Balkans (circa 600 A.D.), but was completed by the time such Christian names as *Laurentius* > *Lovreč* began to be borrowed (sometime after 750 A.D.)

(c) The second palatalization was operative around 600 A.D., was still operative at the time Slavic colonization reached the Adriatic coast, probably ceased by 750 A.D., and certainly ceased before the appearance of the first documents late in the ninth century.

(d) Metathesis of liquids began after *c.* 750 A.D. and was complete before *c.* 900 A.D.

(e) The third palatalization likewise was operative about A.D. 600 and continued to be operative in South Slavic until sometime prior to c. 900.

(f) In the western varieties of South Slavic *y* and *i* early merged into one phoneme *i*; there was probably an accompanying merger of *ŭ* and *ĭ* and a loss of allophonic palatalization before front vowels. This merger took place after monophthongization and after the borrowing of *crucem* > *kryžĭ* and *Rōmam* > **rymŭ* (say sometime after 650–750 A.D.) and before transmittal of these words to the western Slavs (sometime after 850), most probably before the eleventh century.[35]

5.0. Discussions of the development of Slavic contained in the traditional handbooks and other older writings either ignored the problem of the

Slavenstvo i Romanstvo, particularly pp. 35–36 and note 8 on p. 43; also Hamm, Hraste, and Guberina's listing of localities on the island which show cakavism in *Hrvatski Dijalektološki Zbornik* I, 45).

[35] Bulgarian and Macedonian distinguish the reflexes of *ŭ* and *ĭ* (where these reflexes are not zero) and allophonic or phonemic palatalization is retained to the present day in many varieties of Bulgarian, particularly in the East. The falling together of *y* and *i* in the eastern branch of South Slavic must be placed at a later date than that in the western varieties, since OChSl., based on a Macedonian variety of Slavic speech, distinguishes *i* and *y*, at least in the earliest monuments.

chronology of the changes lying between Indo-European and Slavic as we know it from our earliest records, regarding it as more or less unknowable, or else implicitly assigned a very high antiquity to the changes. More recent writers have tended to revise former estimates of the antiquity of these changes (see for example the later writing of N. van Wijk, particularly his series of lectures "Les langues slaves, de l'unité à la pluralité" originally published in *Le monde slave* in 1937), assigning many of the later changes to a time approximating or slightly preceding the first appearance of the Slavs in written history. The net result of this study gives support to the latter view, in that it would appear that many of the later changes were in progress when the Slavs began to expand in various directions around the middle of the first millenium of our era and were completed only shortly before our earliest written records in Slavic.

Department of Modern Languages
University of Pittsburgh
Pittsburgh 13, Pennsylvania

EDWARD STANKIEWICZ

Grammatical Neutralization in Slavic Expressive Forms

1. Of the three functions of language, the cognitive, the appellative and the expressive[1], the last, which renders the attitude of the speaker toward the addressee or the referent, has generated much confusion in linguistic literature. The lack of clarity as to its formal manifestations has even induced some linguists to question the very existence of such a function.

Neoidealistic philologists and some followers of de Saussure have equated the expressive function with the individual creations of "parole," seeing in it the moving force, the "life" of language. According to Charles Bally, the "emotive" is diametrically opposed to the "intellectual," as the subjective is to the objective, the personal to the social, the dynamic to the static.[2] Within such a conception, the linguistic problem could not but dissolve into a stylistics, incapable of distinguishing the expressive invariants of the code from the expressive variants of the message, the primary meaning of lexical items from their secondary emotive overtones, synonyms and free alternants from their stylistic and poetic usage. This approach has, furthermore, magnified the role of the expressive function in the historical development of a language, a role which was never demonstrated to be more than marginal. It should be noted that the imprint of this approach has, nevertheless, been felt in many interpretations of "expressive language" given by prominent structural linguists. Instead of

[1] The three functions of language have been formulated in the well known "Organon-Modell" of K. Bühler's *Sprachtheorie* (1934), pp. 24–33. To these R. Jakobson has recently added three more functions—the poetic, phatic and metalingual; cf. his "Linguistics and Poetics," *Style in Language*, ed. T. A. Sebeok (1960), pp. 353–358. It is questionable whether the latter functions involve linguistic elements other than those which figure in the functions established by Bühler. At any rate, the question deserves further exploration.

[2] *La langue et la vie*, pp. 26–28. Interesting critical comments on this approach are found in L. Hjelmslev's *Principes de Grammaire Générale* (1928), pp. 300 ff.; cf. also his statement (on p. 240): 'Les éléments dits extragrammaticaux ou affectives peuvent en effet obéir aux regles grammaticales, en partie peut-être à des règles grammaticales qu'on n'a pas encore réussi à dégager."

disentangling the various issues of expressive form and function, they have often added to the confusion introduced by scholars of a more idealistic and metaphysical bent.[3]

While languages differ widely in the inventories of expressive (more properly emotive) elements, which may be drawn from various linguistic levels, the isolation and grouping of these elements constitutes only the first step in their analysis. It is only when they are studied with relation to the non-expressive elements that their position within the system as a whole becomes apparent; the expressive elements may, in turn, throw light on the hierarchical organization and the complexity of the various levels of language.

This paper proposes to examine the morphological properties of expressive substantives in the Slavic languages. In traditional Slavic grammars the grammatical peculiarities of the expressive forms have been handled as a conglomeration of scattered, heterogeneous and haphazard facts. The lack of an explanatory principle has made it appear that these forms escape systematization, that they are disturbing "deviants" in the Slavic morphological systems, to be disposed of in lists of exceptions, appendices and footnotes.

2. An over-emphasis on the monolithic nature of language, the insistence on "symmetry" and "overall" patterning, the assumption that over-differentiated sets are matched by homonymous under-differentiated sets, obscure the hierarchical structuring of language, and its correlate, the asymmetry of linguistic relations.[4] Asymmetry involves the neutralization or incompatibility of certain phonological and grammatical oppositions within some sub-systems (paradigmatic neutralization) or in some environments (syntagmatic neutralization). With the emphasis placed on symmetry, which is often found *within* sub-systems, linguists are sometimes prone to forget the asymmetrical relations which obtain *between* sub-systems or between the members of oppositions. Yet, in approaching the

[3] This confusion is quite apparent in the chapter dealing with expressive-emotive elements of V. A. Zvegincev's *Semasiologija* (1957), pp. 167 ff. In the article "Zametki o značenii slova," *Voprosy Jazykoznanija* III (1955), 76, J. Kuryłowicz proposes specifically to subsume expressive forms and stylistic variants under the common heading of stylistics, and proceeds to treat them without distinction later on. According to Trubetzkoy, "expressive phonology" can be compared "a l'étude du costume en ethnographie." His "Phonostylistics" covers "emphatica," variants of social and territorial dialects, and voice-modifiers; cf. *Principes de Phonologie* (1949), pp. 16–29.

[4] Cf. S. Karcevskij, "Du dualisme asymétrique du signe linguistique", *Travaux du cercle linguistique de Prague*, I (1929), 89–93. Despite their sketchiness, Karcevskij's remarks have lost none of their perspicacity and relevance after 30 years.

phonology of an undescribed language, no one expects the vowels to pattern symmetrically with the consonants, or affricates to be matched by an equal number of stops. It is taken for granted that all phonemic systems have empty slots. One of the merits of recent phonemic theory has, in fact, been to show that the combinatory possibilities of the distinctive features, which form phonemic bundles, are not fortuitous, but follow a hierarchical order.[5] While no theory of similar scope and simplicity is available (or perhaps possible) on the morphological level, it is clear that restrictions of a similar type pervade also this level. Thus, it is known that the "marked," more specific, members of grammatical categories are, with respect to their "unmarked" counterparts, less complex, i.e. that they admit fewer combinations with other grammatical categories. For example, the plural of Slavic substantives admits fewer gender and/or case distinctions than does the singular; the preterite of the Russian verb excludes reference to person which is carried by the present; the perfective aspect is incompatible with other marked aspects which are expressed in the imperfective. Functional asymmetry is often accompanied by formal asymmetry, although the form-function relation itself introduces interesting, so far insufficiently explored, asymmetries.[6] Relations parallel to those on the flexional level obtain also on the level of derivation, where the derived forms neutralize distinctions implicit in their underlying simple forms. For example, most Russian derivative adjectives preclude the opposition of short vs. long forms and of grading (i.e. of the positive/comparative /superlative opposition), which are carried by the non-derivative forms; some Russian substantival derivatives do not admit the opposition of number, and most of them have no accentual alternations, the foremost formal, morphophonemic complexity of non-derivative substantives. The asymmetric relationship between primary, basic forms and secondary, "derived" forms cuts thus across flexion as well as derivation, although the types of neutralization are on both axes differently actualized.

3. Neutralization of grammatical and lexical[7] distinctions is one of the essential features of Slavic expressive forms. The types and degrees of

[5] The simplest and most general formulation of this theory is that of R. Jakobson; cf. R. Jakobson, C. Gunnar Fant, M. Halle, *Preliminaries to Speech Analysis. The Distinctive Features and Their Correlates* (1952[1]); R. Jakobson and M. Halle, *Fundamentals of Language* (1956).

[6] I have touched upon this problem recently in "The Consonantal Alternations in the Slavic Declensions," *Word*, XVI (1960).

[7] The discussion of lexical neutralization transcends the scope of this paper. Suffice it to point out that expressive derivation is an unexhausted source of synonyms, i.e. of words with an identical referential meaning but with different stylistic or expressive

neutralization in the Slavic languages form a continuous scale which is correlated with the various degrees of emotive content rendered by expressive forms. Between the degrees of intensity of emotion and degrees of "grammaticalness" there is, in fact, an inverse relationship. The most expressive elements in language, the simple interjections (like the Russ. *ax, jej, fu, t'fu*), have the lowest degree of lexical specificity, and are morphologically the least characterized elements of the system. Syntactically they function as autonomous "phrases non-conceptuelles."[8]

The interjectional verbs (e.g. Russ. *pryg, tolk, bux*) or nouns (e.g. Russ. *góspodi, mámon'ki, bóže*) contain, on the other hand, a lower expressive charge, but a higher degree of lexical and grammatical (morphological–syntactic) specificity. As is known, interjections sometimes also differ phonemically from non-expressive forms: they include phonemes and phonemic sequences (such as the Russ. /γ/ in *góspodi*, or /t'f/ in *t'fu*) which are also found in quotation-words or onomatopoeia, and which can be interpreted as a peripheral sub-system or placed outside the "normal" phonemic system (the customary solution). Given the hierarchical and multi-class nature of morphological systems, even the most expressive forms can be treated rather within than outside the system, exhibiting simply different degrees of grammatical complexity.

4. The morphological "irregularities" of the Slavic expressive substantives form a pattern, or patterns, which must be examined both with relation to the grammatical structure of the non-expressive forms and in terms of the lexical structure of the expressive forms themselves.

Before discussing grammatical neutralization in expressive substantives, we shall define the criteria for their identification and classification.

The Slavic expressive substantives can be grouped into three types, according to the structure of their stems and to their expressive function.

4.1. The first type is that of *substantive derivatives*. They comprise substantive stems which are followed by diminutive/augmentative and/or affectionate/pejorative suffixes. The class membership and lexical meaning

overtones. This will be apparent in reading this paper, where many words are given an identical translation into English, lacking equivalent stylistic or expressive nuances; cf. the Russ. *p'jánica, propójca, p'jančúga*, Pol. *moczymorda, moczygęba, moczybroda, moczywąs, płóczywąs, kropiwąs, dusipiwo, morzypiwo, pijus, pijak, opijus, opilczuch*, etc., all of which mean 'drunkard'. A treasure-house of Polish synonyms with this meaning is J. Tuwim's *Słownik pijacki* (1930).

[8] Cf. S. Karcevskij, "Introduction a l'étude de l'interjection," *Cahiers F. deSaussure*, I (1941), 57; J. M. Kořinek, "Zur lautlichen Struktur der interjektionalen Sprachgebilde," *Slavia*, XV (1937–38), 43 ff.; G. V. Dagurov, "O leksičeskom sostave russkix meždometij," *Slavjanskoe Jazykoznanie* (1959), pp. 154 ff.

of the basic forms remains intact in these derivatives, although their inventories and functions may vary in the individual languages. Examples of such forms are: Russ. *synók* (*syn* 'son'), *mátuška* (*mat'* 'mother'), *réčen'ka* (*reká* 'river'), *toporíšče* (*topór* 'axe'), *domíško* (*dom* 'house'); Ukr. *rúčka* (*ruká* 'hand'), *didús'* (*did* 'grandfather'), *babúsja* (*bábka* 'grandmother'), *xlopýšče* (*xlópec'* 'boy'); Pol. *kotek* (*kot* 'cat'), *dziewczątko* (*dziewczyna* 'girl'), *babsko* (*baba* 'woman'), *kobiecina* (*kobieta* 'woman'); Cz. *větérek* (*vítr* 'wind'), *chvilečka* (*chvíle* 'moment'), *psíček* (*pes* 'dog'), *chlapisko* (*chlap* 'fellow'); S-Cr. *brâtac* (*brât* 'brother'), *stvárca* (*stvâr* 'thing'), *jèzičina* (*jèzik* 'tongue'), *knjižùrina* (*knjìga* 'book'), etc.[9]

To this type belong also hypocoristic formations, in which the emotive meaning is rendered through truncation of the stem with or without suffixation. Hypocoristica are especially common in proper names, which function sometimes not as emotive forms but as nicknames, but pattern formally like the derivatives of common substantives; they are, like these, based on full, official forms of names, and are subject to grammatical neutralization.

In some Slavic languages (e.g. S-Cr.), hypocoristic formation is productive also in common substantives, particularly in nursery words referring to immediate, intimate relatives; the hypocoristica may be used with or without expressive suffixes. Examples of hypocoristica (of proper names) are: Russ. *Tánja* (*Tat'jána*), *Šúra* (*Aleksándr*, *Aleksándra*); Cz. *Pepa* (*Josef* or *Josefa*); S-Cr. *Ráde* (*Rádslav*), *Pájo* (*Pâvao*); Slovak *Jano* (*Ján*), *Vládo* (*Vladimir*); Pol. *Stefcio* (*Stefan*), *Józio* (*Józef*); Ukr. *Stepánko* (*Stepán*), Bulg. *Kól'o* (*Nikóla*); (of common substantives): S-Cr. *zéko* (*zêc* 'rabbit'), *brále* (*brât* 'brother'), *príko* or *príšo* (*prijatelj* 'friend'); Slovak *zajko* (*zajac* 'rabbit'), *byco* (*býk* 'bull').

4.2. The second type of expressive formations consists of *substantivized derivatives*. These are built almost exclusively on verbal and adjectival stems. They are found primarily in colloquial and substandard language, and carry the meaning of irony, contempt, and less frequently of affection. The reification of attributes resulting from the transformation of a modifier into a substantive makes these forms particularly suitable to function as highly emotive, "loaded" curse words or nicknames for human beings (Russ. *kličkí*). Many of them attach derivative suffixes. Examples (without

[9] All forms quoted in this paper are transliterated. Expressive suffixes and desinences are marked by a preceding hyphen, but are, for the sake of clarity, not given in consistent morphophonemic transcription when this is irrelevant to the central problems of the exposition. The English translations are, for obvious reasons, only approximate equivalents of the Slavic expressive forms. In the case of substantive derivatives, the translation is of the basic forms only.

suffixes) are: Russ. *neposéda* 'fidget', *obžóra* 'glutton', *zaíka* 'stammerer', *povésa* 'rake'; Pol. *niezdara* 'clumsy', *kaleka* 'cripple'; S-Cr. *ćóro* 'one-eyed', *ćélo* 'bald-headed', *šánto* 'lame', *gúro* 'hunchback', *gól'o* 'poor fellow'; Cz. *necuda* 'shameless person', *nejapa* 'fool', *nezdara* 'clumsy', *dareba* 'rogue', *zámora* 'weak, poor fellow', *bačkora* 'coward'; (with suffixes): Russ. *gromíla* 'burglar', *kutíla* 'rake', *guljáka* 'idler', *lakómka* '*gourmand*', *deljága*, *rabotjága* 'industrious one', *toropýga* 'fidget', *projdóxa* 'cunning one', *pláksa*, *xnýksa* 'sniveller', *propójca*, *p'jančúga* 'drunkard', *tupíca* 'dumbbell'; Pol. *jąkała* 'stutterer', *bazgroła* 'scribbler', *włóczęga* 'vagabond', *niedołęga* 'clumsy', *beksa*, *płaksa* 'blubberer', *grubas* 'fatso', *brudas* 'dirty', *śpioch* 'sleepy-head', *leniuch* 'lazybones', *głupek* 'fool', *obdartus* 'ragged'; Ukr. *zabúd'ko* 'forgetful person', *neznájko* 'ignoramus', *harýkalo* 'grumbler', *xvol'kó* 'braggart'; S-Cr. *blebètalo* 'gossip', *zadirkívalo* 'tease', *drémalo* 'sleepy-head', *pripovédalo* 'chatter-box'; Bulg. *glúpčo* 'fool', *krádl'o* 'thief', *lážl'o* 'liar', *bəbrívko* 'chatterer'; Slovak *trul'o* 'blockhead', *chmul'o* 'gawk', *chlipko* 'drunkard', *bojko* 'timid fellow' Maced. *plačko* 'cry-baby', *mrzlo* 'lazybones', *partalko* 'beggar'.[10]

To this same group belong also pejorative, ironical nickname-type compounds which are transforms of elementary phrases (syntagms), consisting in most cases of verbal plus nominal constituents; e.g. Russ. *ležebóka* 'lazybones', *pustomélja* 'chatterer', *deržimórda* 'boor', *sorvigolová* 'mad-cap', *krivošéjka* 'crooked-necked person'; Pol. *zawalidroga* 'lubber', *moczymorda* 'drunkard', *wiercipięta* 'giddy-head', *pasibrzuch* 'glutton', *półgłówek* 'half-wit'; Slovak *ludigroš* 'miser', *prepiduša* 'drunkard', *potrimiska* 'parasite'; S-Cr. *ràspikuća* 'spendthrift', *pŕžibaba* 'robber (abuser of women)', *mòlibog* 'monk' (ironical), *pläčidrūg* 'friend' (in trouble, in tears); Bulg. *nexranimajko* 'rascal', *varíklečko* 'miser', *lapnímuxa* 'a fool'; Maced. *lapnikandilo* 'glutton', *zadripetle* 'thief'.

4.3. The third type of expressive forms consists of *simple, non-derivative substantives*. Their expressive meaning, affectionate or pejorative, is inherent in the root-morpheme. The affectionate forms belonging to this type are almost exclusively nursery-appellations for kinship-terms, similar to the hypocoristica of type I, but with full stems which may or may not be followed by expressive suffixes. They have been characterized as intermediate between common and proper nouns (*polunaricatelni, naricatelno-sobstveni imena*[11]). As opposed to the adult, more abstract and emotionally

[10] In derivatives of types I and II, the various Slavic languages employ different sets of expressive suffixes. A discussion of them is beyond the scope of this paper. For a description of the expressive suffixes of type I, cf. the author's "Expressive Derivation of Substantives in Contemporary Russian and Polish," *Word* X (1954), 457 ff.

[11] E. Georgieva, "Mama i majka," *B"lgarski ezik*, IX (1959), 287 ff.

'neutral' kinship-terms, they carry intimate, endearing, childishly tinged connotations; one of their characteristic morphological features is the nom. sg. *-o* desinence for masculine substantives, which functions simultaneously as an expressive signal (cf. 6.1). Examples are: Ukr. *dído* (*did* 'grandfather'), *táto* 'father'; Pol. *wujo* (*wuj* 'uncle'), *tato* 'father'; S-Cr. *bábo*, *ćáko* 'father', *dédo* 'grandfather'; Bulg. *bájo* (*baj*) 'older brother, person'.

The pejorative forms (contemptuous or ironical) are likewise used primarily as appellatives, like proper nouns; e.g. Russ. *šél'ma* 'rogue', *kanál'ja* 'rascal', *skrjága* 'miser', *xanžá* 'hypocrite'; Pol. *gbur*, *cham* 'crude fellow', *gach* 'lover, seducer', *dziad* 'beggar' (homonym of *dziad* 'grandfather'). The criterion for setting these forms apart from other substantives which carry pejorative connotations lies precisely in their grammatical treatment, i.e. in their being subject to neutralization, like the derivatives of types I and II. It is here suggested that these forms must be clearly distinguished from substantives in which the expressive function arises only secondarily, i.e. in metaphoric usage. Expressive metaphors owe their meaning to verbal or situational contexts, but are not systemically expressive. The lack of distinction between the former and the latter (another case of the confusion of message with code) inevitably obscures the linguistic aspect of 'expressiveness' and leads to its confusion with stylistics. The examples of 'expressive forms' quoted in Russian works, such as *bába*, *mókraja kúrica* 'female, wet chicken', or *sobáka* 'dog' (a form of endearment used by Chekhov in letters to his wife), illustrate only the expressive possibilities of the message; outside of the verbal or situational contexts or of private metaphoric usage, these forms are not expressive. The same holds for emotive switching of gender, which has been observed in many languages (cf. the Russ. *dúra* 'fool' (f.) applied to a male, Ukr. *sýne* 'son' to a daughter, French *ma vieille* with reference to a man, or *mon petit* applied to a wife).[12]

Placed within the proper context, almost any word can be given an expressive value, which does not prevent languages from conventionalizing some of them as metaphoric expressive devices *per excellence*. Favorite metaphoric, in particular pejorative appellatives are, for examples, names

[12] In his discussion of grammatical gender, in *Russkij jazyk* (1947), pp. 73 ff., Vinogradov cites older linguistic literature and interesting examples from Russian literature on the expressive use of feminine gender. The stylistic and linguistic aspects merge almost imperceptibly in Vinogradov's treatment. The same unfortunate fusion takes place in F. Oberpfalcer's "Rod jmen v češtině," *Charisteria G. Mathesio oblata* (1932), pp. 85–91. Examples of expressive switching of gender from various European languages are given by St. Wędkiewicz in "Kilka uwag o technice spieszczeń," *Język Polski*, XIV (1929), 110 ff.

of animals, of plants, of parts of the body (e.g. Eng. *pig*, *snake*, *louse*, *rat*; Russ. *zmejá* 'snake', *svin'já* 'pig', *lisá* 'fox', *sobáka* 'dog', etc.); items of clothing may also enjoy a favorite position, such as Russ. *šljápa* 'hat', *trjápka* 'rag', or even musical instruments, such as Pol. *trąba* 'trumpet', *fujara* 'fife', *bęben* 'drum' (figuratively they all mean 'clumsy person', 'fool').

The synchronic difference between contextually expressive variants and expressive invariants by no means precludes their dynamic interaction. When the expressive meaning of the variants acquires an autonomous value, we are faced with a diachronic split, i.e. with homonyms, which differ from each other both lexically (in their primary, basic meanings) and grammatically. Such homonyms are, for example, the Russian *stérva* (f.) 'carrion' vs. *stérva* 'vile person', *dubína* (f.) 'heavy stick' vs. *dubína* 'stupid fellow', *skotína* (f.) 'cattle' vs. *skotína* 'crude fellow', *baldá* (f.) 'heavy hammer' vs. *baldá* 'useless fellow'. The first forms are of feminine gender, whereas the latter are so-called 'epicoena' (of 'common gender'). Similar homonyms are the Pol. *galgan* 'rag' vs. *galgan* 'rogue', *cymbały* 'cymbals' vs. *cymbał*, pl. *cymbały* 'simpleton', *bydlak* 'beast' vs. *bydlak* 'vulgar person'. The first members of the pairs are grammatically inanimate, whereas the second are animate-personal substantives. And like the other expressive substantives, they are subject to grammatical neutralization.

5. Grammatical neutralization in expressive substantives involves both the suppression of *functional* oppositions (i.e. of number, gender and case) and/or of *formal* distinctions (of particular desinential suffixes and declensional paradigms). Some types of neutralization are conditioned by and predictable from the very structure of the expressive substantives. Thus we may further distinguish (1) *syntagmatic neutralization* pertaining to substantives of type I; it involves the suppression of grammatical distinctions inherent in the underlying, simple forms from which these substantives are derived, and (2) *paradigmatic neutralization*, cutting across all three types of expressive substantives; it involves the elimination of oppositions inherent in the nominal system of non-expressive forms.

We shall examine in turn the neutralization of functional and then of formal oppositions.

5.1. Expressive substantives exhibit in all Slavic languages neutralization of the opposition of number. This neutralization is the more common the higher the emotive charge of the expressive form. Thus it occurs primarily in derivatives with an augmentative-pejorative or with a strong affectionate meaning. For example, Russian augmentatives with the suffix *-išče* and

affectionate diminutives with the complex suffix *-iška*, Macedonian diminutives with the suffixes *-ec*, *-ence*, Polish diminutives with the suffixes *-inka*, *-ulek*, *-uńko*, *-eńko* rarely admit the plural.[13] Only a small number of Russian augmentatives (e.g. *družíšči* 'friends', *syníšči* 'sons', *domíšči* 'houses') may occur in the plural, whereas a few derivatives (with a somewhat modified lexical meaning) occur only in the singular (*delíški* 'business', *kartíški* 'cards', *svjazíški* 'connections'). Within the Slavic nominal systems, expressive derivatives behave thus, in distinction to their underlying non-expressive forms, like *singularia tantum*, or in limited instances like *pluralia tantum*.

5.2. Far more complicated than neutralization of number is neutralization of gender distinctions in expressive forms, inasmuch as the treatment of gender itself varies significantly from one Slavic language to another. The category of gender is, in the Slavic languages, correlated with the singular/plural dichotomy; as stated above, the singular carries more and/or different gender distinctions than the plural. In the singular, which is in this respect homogeneous in all Slavic languages, substantives belong to the masculine animate or inanimate, feminine, or neuter gender. The grammatical genders carry differently marked meanings with respect to "natural gender," i.e. with respect to the sex of animate beings: the masculine gender may designate members of either sex, the feminine cannot designate a male, and the neuter cannot designate a male or female (in Russian it refers, with a few exceptions, to inanimate beings, in other Slavic languages it also includes the young of animals and of human beings).[14] The plural of substantives expresses, on the other hand, in most Slavic languages (except Serbo-Croatian, Slovenian, and Standard Czech) the opposition between a personal/impersonal gender (Polish, Slovak, Lusatian), or between an animate/inanimate gender (Russian and, with some variations, Ukrainian and Byelorussian), or altogether lacks gender (Bulgarian, Macedonian). The rules for gender of expressive forms thus ultimately depend, like those of non-expressive forms, on the dichotomy of number. Syntagmatic neutralization of gender, i.e. the overlap of genders

[13] Slavic grammars only occasionally register neutralization of number in expressive derivatives; but cf. M. Rešetar, *Elementargrammatik der serbischen (kroatischen) Sprache* (1946), p. 46; B. Koneski, *Gramatika na makedonskiot literaturen jazik*, II (1954), 14, 39.

[14] It is noteworthy that masculine substantives admit feminine modifiers when they refer contextually to a woman and have expressive meaning; e.g. Cz. *kluk zlodějská*, *lotr lotrovská*, *lump lumpácká*; cf. Oberpfalcer (op. cit.); the same possibility exists also in non-expressive phrases, when the gender is specified referentially; e.g. Russ. *vrač'* or *sekretar' vyšla*.

in derivatives from stems in which they are distinctive, operates in expressive substantives of type I, while paradigmatic neutralization, i.e. the complete elimination of gender, affects expressive substantives of the other types, including some hypocoristica of proper names (of type I).

5.21. Neutralization of the first kind occurs when substantives of masculine, feminine, or neuter gender exhibit in their expressive derivatives only one gender, specifically the neuter, and, less commonly, the feminine. Thus augmentative–pejorative derivatives are in various Slavic languages of neuter gender, whereas their underlying simple forms are of the masculine, feminine or neuter gender; e.g. Pol. *chłop* 'peasant, chap', *kot* 'cat' (m.), *kobieta* 'woman', *krowa* 'cow' (f.), *ciało* 'body' (n.) vs. *chłopisko, kocisko, kobiecisko, krówsko, cielsko* (n.); Slovak *chlap* 'man, fellow' (m.), *kniha* 'book' (f.), *mäso* 'meat' (n.) vs. *chlapisko, knižisko, mäsisko*; Bulg. *mъž* 'husband' (m.), *žená* 'wife' (f.), *kúče* 'dog' (n.) vs. *mъžíšte, ženíšte, kučétište* (n.).

A similar neutralization is bound up in Polish with the affectionate suffix *-ątko*, in Slovak with *-atko*, in Bulgarian with the suffixes *-če/-iče, -ence/-čence*, or with the pejorative suffix *-le*.

Overlapping of the masculine with the feminine gender is limited to derivatives with the suffix *-ina* (and its compound extensions, such as S-Cr. *-èt-ina, -úr-ina, -ùš-ina, -č-ina, -ùšt-ina*); in various Slavic languages the suffix has an affectionate or pejorative-augmentative meaning, e.g. Pol. *pies* 'dog' (m.) *noga* 'leg' (f.), *drzewo* 'tree' (n.) vs. *psina, nożyna, drzewina* (f.); Slovak *dub* (m.) 'oak', *vŕba* (f.) 'willow' vs. *dubina, vŕbina* (f.); S-Cr. *òrao* 'eagle', *jùnāk* 'brave man, hero', *òganj* 'fire' (m.) *rúka* 'hand', *žéna* 'wife, woman', *vàtra* 'fire' (f.) vs. *òrlina* (or *orlètina/orlùština*), *junàčina ognjùština, rùčina* (or *ručètina/ručùrina*), *ženètina,vatrùština* (f.). Masculine substantives attaching the suffix *-ina* become feminine in Russian only when they are inanimate; e.g. *xvost* 'tail', *gólos* 'voice' vs. *xvostína, golosína*, while animate substantives retain the gender of the basic stem, e.g. *starík* 'old man', *kupéc* 'merchant', *zver'* 'animal' (m.) vs. *staričína, kupčína, zverína* (m.).

The same system sometimes reveals a state of oscillation between neutralization and preservation of the inherent gender. Thus animate masculine substantives with the suffixes *-isko* and *-ina* in Polish are mainly neuter or feminine respectively, but they may also retain the gender of the basic forms; e.g. *człeczysko, chłopaczysko, kocisko, konisko*; *psina* (derived from *człek* 'man', *chłopak* 'boy', *kot* 'cat', *koń* 'horse', *pies* 'dog'). Russian presents free variation in a few derivatives of animate substantives in *-ina*; e.g. *volčína, psína, idiótina* (derived from *volk* 'wolf', *pës* 'dog',

idiót 'idiot'). What are 'free variants' within one and the same system, can appear also as different solutions between various systems, particularly between dialects. In Standard Slovak (which is based on the Central dialect) gender is neutralized in derivatives with *-isko/-sko*; e.g. *chlapisko jazyčisko, medvedisko*—from *chlap* 'fellow', *jazyk* 'tongue, language', *medved'* 'bear' (m.); in Eastern and Western Slovak, however, the basic stem gender remains intact.[15]

5.22. Suppression of all gender distinctions in the singular is found in expressive substantives of types II and III and in some hypocoristica. Traditional grammar calls these forms without gender 'epicoena.' But the gender-category called 'epicoena' (i.e. of common gender) is a *contradictio in adjecto*, because the distinctive property of substantives as opposed to adjectives with their variable gender is in the Slavic languages (as in some other languages) their invariant, inherent gender.[16] Since all 'epicoena' designate human beings, their "variable," masculine or feminine gender is determined extra-linguistically, i.e. through the sex of the referent which is either 'male' or 'female'. In the languages which have a case system, the 'epicoena' belong to the so-called '*a* declension', which encompasses both masculine and feminine substantives; it is only in the plural that the animate or personal gender of the 'epicoena' is expressed grammatically. All Slavic languages offer examples of expressive substantives of the 'common gender'. They vary merely as to the degree of productivity of this category, which includes substantivized derivatives, inherent expressive words, and many hypocoristica of proper names. Examples are: Russ. *kutíla* 'rake', *nevéža* 'ill-mannered person', *sónja* 'sleepy-head', *bednjága* 'poor fellow', *slastëna* 'sweet tooth', *deržimórda* 'boor', *skotína* 'brute', *lapšá* 'muddled person', *baldá* 'blockhead'; Pol. *włóczęga* 'vagabond', *beksa* 'sniveller', *gderała* 'grumbler', *wiercipięta* 'giddy-head'; Cz. *dareba* 'rogue', *chudera* (or *chuděra*) 'poor creature', *zámora* 'weakling'; S-Cr. *konjòmora* 'reckless rider', *pàlikuća* 'arsonist'; Bulg. *pijánica* 'drunkard', *ubíjca* 'murderer'; Maced. *partalka* 'beggar', *plačko* 'cry-baby' (for additional examples, cf. **4.2**); examples of proper names are: Russ. *Pólja* (*Apollónij, Polikárp* and *Polína, Pelagéja*), *Tólja* (*Anatólij* and *Kapetolína*), *Lëlja* (*Lev* and *Eléna*); Pol. *Leszuchna* (*Leszek* and *Leszka*); Cz. *Tona*

[15] Cf. A. Jánošík, "Gramatický a prirodzený rod augmentatív na-isko," *Slovenska Reč*, XIII (1947–48), 17 ff.

[16] The gender category called 'epicoena' is also rejected by A. N. Gvozdev, *Sovremennyj russkij literaturnyi jazyk*, I (1958), 138; the epicoena, says Gvozdev, are "imprecisely called words of common gender; their gender varies according to each concrete case, in which it refers to a man or a woman."

(*Antonín* and *Antonína*), *Pepa*, *Pepička* (*Josef* and *Josefa*, *Josefina*), *Míla* (*Miloslav* and *Miloslava*, *Milena*).

Not all expressive derivatives of type II are epicoena. In some languages they are rather of masculine gender, this gender permitting reference to both males and females. Such, for instance, are the Pol. *leniuch* 'lazybones', *brudas* 'slob'; Cz. *miláček* 'sweetheart', *morous* 'grumbler', *chudák* 'poor fellow'; Bulg. *varíklečko* 'miser', *lapnímuxa* 'fool'. Some expressive substantives, on the other hand, oscillate between the masculine gender and the absence of gender; e.g. Russ. *podlipálo* (m.) or *podlipála* (m. or f.) 'flatterer'; Bulg. *nexranímajko* (m.) *nexranímajka* (m. or f.) 'rascal', *podlizúrko* (m.) or *podlizúrka* (m. or f.) 'flatterer'; Maced. *mrzlo* (m.) or *mrzla* (m. or f.) 'lazybones', *partalko* (m.) or *partalka* (m. or f.) 'beggar'.

5.23. As opposed to the singular, the plural of expressive substantives (if it is used at all) admits at best only partial neutralization of gender. This seeming restriction in the occurrence of neutralization is, however, explainable if we remember that in most Slavic languages the plural of non-expressive substantives has only the opposition animate/inanimate or personal/impersonal, or has no gender at all. The animate or personal gender of the hypocoristica and expressive forms of types II and III (which refer to human beings and some animals) is expressed in the accusative in the languages with case systems, and also in the nominative plural in some of them. Partial neutralization of the personal/impersonal gender is found in the plural of two Slavic languages—Slovak and Polish. In Slovak it is limited to a few terms for animals, which are of the personal gender in the simple forms, but which are depersonalized in the derivative forms. Thus the neutral forms *vtáci*, *býci*, *vlci*, *psi* are personal (from *vták* 'bird', *býk* 'bull', *vlk* 'wolf', *pes* 'dog'), but their diminutives *vtáčiky*, *býčky*, *vlčky*, *psíky* are impersonal.[17]

In Polish, where non-expressive forms render the personal/impersonal opposition both in the accusative and in the nominative plural (cf. nom. pl. *koledzy*, *doktorzy*, *panowie*, acc. pl. of personal nouns *kolegów*, *doktorów*, *panów*, vs. the nom.-acc. pl. of impersonal nouns *koty*, *stoły*, *kobiety*, *okna*), the expressive forms of types II and III neutralize the gender opposition in the nominative plural; e.g. *włóczęgi* (*włóczęga* 'vagabond'), *młodziaki* (*młodziak* 'youngster'), *brudasy* (*brudas* 'sloven'), *żarłoki*, *pasibrzuchy* (*żarłok*, *pasibrzuch* 'glutton'), *moczymordy* (*moczymorda* 'drunkard'), *gbury* (*gbur* 'boor'), *chamy* (*cham* 'peasant, churl'), *gałgany* (*gałgan* 'rogue'), but the acc. pl. is *włóczęgów*, *młodziaków*, etc. The personal/impersonal opposition is likewise eliminated in the nominative plural of

[17] E. Pauliny, J. Štolc, J. Ružička, *Slovenská gramatika* (1955), p. 154.

Polish derivatives in *-isko* or *-ina*; e.g. *chłopczyska, człeczyny* (but acc. pl. *chłopczysków, człeczyn*).

The connection between grammatical neutralization of gender and expressive language is most clearly illustrated by Bulgarian and Macedonian, where it has, however, received a different solution. In these two Slavic languages, which have no gender oppositions in the plural, the difference between an emotive vs. non-emotive attitude is itself grammaticalized in the plural of *non-expressive* substantives. Thus Bulgarian simple masculine substantives can in the plural be opposed as "neutral" vs. expressive (pejorative) forms; e.g. *drúmove* vs. *drúmišta* (from *drum* 'road'), *gospodá* vs. *gospodínovci* (from *gospodín* 'mister, man'), while some neuter substantives are in Macedonian opposed as "neutral" vs. expressive (affectionate) forms; e.g. *srca* vs. *srcina* (from *srce* 'heart').

A similar opposition between expressive vs. non-expressive pejorative forms is available also in the nominative plural of Polish masculine non-expressive substantives; e.g. *doktorzy* vs. *doktory* (*doktór* 'doctor'), *Żydzi* vs. *Żydy* (*Żyd* 'Jew'), *rzeźnicy* vs. *rzeźniki* (*rzeźnik* 'butcher').[18]

The grammaticalization of emotive attitudes in a morphological category which has no or few gender distinctions is but the reverse aspect of the degrammaticalization of gender distinctions in lexical forms (primarily derivative) which carry emotive meanings. Both phenomena point up the close interdependence between the grammatical and emotive functions of language.[19]

5.3. Neutralization of case distinctions is found only in the most expressive forms, in the hypocoristica. As it is connected with the elimination of formal distinctions and with the liquidation of declension, it will be discussed below (cf. § 6.2).

6. Neutralization of formal distinctions refers to discrepancies between expressive and non-expressive substantives in their desinential suffixes and declension. These discrepancies are found in the singular, particularly in the nominative and in the vocative singular (the latter, as a form of address, is outside the case system).

In order to discuss the phenomena of formal neutralization, it is necessary to consider first the relationship of Slavic declensional paradigms to

[18] Emotive grammaticalization in the plural is discussed in greater detail in my forthcoming "The Singular/Plural Opposition in the Slavic Languages," *International Journal of Slavic Linguistics and Poetics*, IV.

[19] This interdependence is manifested also in the Slavic adjectives. The Russian adjectives, for example, carry either expressive or grammatical distinctions (gradation, short vs. long forms). The combination of the two functions is incompatible.

gender and the capacity of the substantive suffixes to specify gender. The following facts are pertinent to our problem. In the Slavic languages, the so-called '*-a* declension' is non-neuter, whereas the so-called '*-o* declension' is non-feminine. The latter declension includes masculine substantives which have in the nominative (-accusative) a zero desinence, and neuters which have in the same cases an *-o* desinence. In some Slavic languages the two genders are set apart through different suffixes of the oblique cases as well; e.g. Pol. (gen. n.) *miasta* 'city', *serca* 'heart' vs. (gen. m.) *ogrodu* 'garden', *stołu* 'table'; (dat. n.) *miastu*, *sercu* vs. (dat. m.) *ogrodowi*, *stołowi*. Some languages, furthermore, show a tendency to differentiate the animate and inanimate gender in cases other than the accusative singular; e.g. Cz. (gen. anim.) *bratra* 'brother', *studenta* 'student' vs. (gen. inanim.) *stolu* 'table', *zvonu* 'bell'; (dat.-loc. anim.) *bratrovi*, *studentovi* vs. (dat. inanim.) *stolu*, *zvonu* (loc. inanim.) *stole*, *zvonu*.

6.1. The declension of expressive substantives exhibits forms which contradict the above rules. Specifically, masculine expressive substantives show up in the nominative singular with the ending *-o*. This ending (which is otherwise a marker of neuter substantives) obliterates the formal distinction between neuter and masculine substantives completely or only partially, depending on the lack or presence of desinential differences in the oblique cases (including the difference between animate/inanimate substantives in the accusative singular). The suppression of formal distinctions in the nominative and in the vocative is the more important in that expressive words are primarily used in these two forms, of naming and address, and less commonly in the oblique cases. Thus even if the masculine or neuter gender of expressive substantives remains intact, they are not distinguished morphologically (as are the non-expressive forms), but, at best, syntactically. The extent of this neutralization varies from one Slavic language to another, and according to expressive type. In Eastern Serbia (especially around Belgrade) and in Czech, it is limited in all three types. In Russian it occurs only in masculine derivatives of type I with the suffixes *-išče*, *-iško* (with the latter only in inanimate substantives); e.g. *syníšče*, *gorodíšče*; *rotíško*, *domíško* (from *syn* 'son', *górod* 'city', *rot* 'mouth', *dom* 'house'). In the other Slavic languages (or dialects), it is found in all types: e.g. (I) Ukr. *sýnko*, *vújko*, *Márko*, *didýšče*/*didún'o*/*dídon'ko*/*didúsen'ko*, *baranýšče*, *bábýšče*, *sokolón'ko* (from *syn* 'son', *vuj* 'uncle', *did* 'grandfather', *barán* 'ram', *bába* 'woman', *sókil* 'falcon'); Pol. *ojcunio*, *dziadunio*, *wujcio*, *syncio*, *Józio*, *Władzio* (from *ojciec* 'father', *dziad* 'grandfather', *wuj* 'uncle', *syn* 'son', *Józef*, *Władysław*); Slovak *zajo*, *býčo*, *otecko*, *oco*,

strýko, Jano, Emo (from *zajac* 'rabbit', *býk* 'bull', *otec* 'father', *Ján, Emil*); S-Cr. (South and West) *bráco, póbro, cígo, Bóžo, Táde* (from *brȁt* 'brother', *pòbratim* 'sworn brother', *cìganin* 'gypsy', *Bóžidar, Tádija*); Bulg. *sínko, tátko, Kól'o* (*sin* 'son', *tato* 'father', *Nikóla*); Cz. *Jirko, Mirko* (*Jiří, Miroslav*). (II) S-Cr. *zadirkívalo* 'joker', *blebétalo* 'chatterer', *góljo* 'poor wretch', *drénjo* 'lazybones', *gúro* 'hunchback'; Ukr. *zdorovýlo* 'healthy fellow', *neznájko* 'ignoramus'; Pol. *jękajlo* (or *jękała*) 'stammerer'; Slovak *chlipko* 'drunkard', *bojko* 'timid fellow'; Bulg. *glúpčo* 'fool', *drípl'o* 'tramp', *nexranímajko* 'scoundrel'; Maced. *plačko* 'cry-baby', *mrzlo* 'lazybones', *partalko* 'beggar'. (III) Pol. *tato* 'father', *wujo* 'uncle'; Slovak *tato* 'father', *dedo* 'grandfather'; Ukr. *dido, tato/táten'ko/tatún'o*; S-Cr. (South and West) *bábo, čáko* 'father', *dédo* 'grandfather', Bulg. *táto/tátko* 'father', *bájo/bájko/bájčo* 'older brother or person'.

As noted above, forms in *-o* (of the non-feminine declension) alternate with forms with *-a* (of the non-neuter declension) either along dialectal lines (cf. the Eastern vs. the South-Western Serbo-Croatian varieties), or according to type. On the whole, masculine expressive forms in *-o* prevail in the South Slavic languages and in Ukrainian, whereas in the North Slavic languages, particularly in standard Russian, Czech, and Polish, they are matched by 'epicoena' in *-a* or by forms of the masculine declension with a zero ending in the nom. sing.

The expressive substantives in *-o* also show another peculiarity: they employ *-o* and *-e*, which are in non-expressive forms in complementary distribution either as free variants, or they generalize one variant independently of the quality of the final stem consonant; e.g. Maced. *tatko* or *tate* 'father'; S-Cr. *Máto* or *Máte* and *príšo, újo, brájo* (*Màtija, prijatel* 'friend', *ŭjak* 'uncle', *brȁt* 'brother'), side by side with *kâle, brâle, Ráde* (*kàludjer* 'monk', *brat, Radoslav*); Bulg. *brátko* or *bráte, bájo* or *báe* (*brat* 'brother', *baj* 'older brother or person'). This phenomenon is particularly common in the vocative.

6.2. Formal distinctions are maximally neutralized when the difference between feminine and masculine-neuter endings is eliminated. This takes place when feminine substantives too take the *-o/-e* ending in the nominative singular, even though they are otherwise declined according to the *-a* declension. When the masculine forms too fall into this declension, as they do in Eastern Serbo-Croatian dialects, the overlap of their flexion is complete: e.g. S-Cr. (m.) *brâle* (*brȁt* 'brother'), *tâle* (*táta* 'father'), *čîle* (*čîdâš* 'horse'), *Krîle* (*Křsta*), *Îve* (*Îvan*), and on the other hand (f.) *púne* (*pùnica* 'mother-in-law'), *néve* (*nèvesta* 'bride'), *bábe* (*bàba* 'grandmother'), *séle* (*sèstra* 'sister'), *Máre* (*Marija*). But in actuality these hypocoristica are

rarely declined since they are primarily used as appellatives, i.e. in the vocative (where they have a falling accent, as opposed to the other forms whose typical prosodic feature is the rising accent). Neutralization of formal distinctions is also found in masculine and feminine derivatives which take on a zero ending; e.g. the Polish derivatives of feminine and masculine proper names, such as *Janek*, *Janúś* (*Jan*), *Juliś* (*Juliusz* or *Julian*), *Zosik* (*Zofia*), *Jadzik* (*Jadwiga*). These hypocoristica lose not only the gender of their underlying forms, but also their flexion. In a Polish phrase such as *nie bierzcie moje Hanuś na pochów*, the derivative *Hanuś*, which in its base form is feminine, is of neuter gender and is treated as an undeclinable.[20] The example also illustrates that the function of such forms is not necessarily limited to that of a subject or a form of address; in the latter functions they merely have the highest frequency of occurrence,

7. Complete neutralization of grammatical distinctions is the rule in the vocative, in which the combination of the appellative and emotive functions coincides with the suppression of all grammatical oppositions characteristic of the substantive. Gender distinctions are obliterated through the use of only one ending for masculine and feminine nouns, or through the complete elimination of the desinence and the reduction of the expressive form to a bare stem. The latter treatment presents the neutralization of all grammatical distinctions.

The generalization of one desinence, including that of the variants of one desinence, is found in the vocative of various expressive forms, but especially in the highly emotive nursery words and proper names. The vocative endings preferred in Polish and Ukrainian expressive substantives is *-u*; e.g. Pol. (m.) *tatulu* (*tatulo* 'father'), *dziadziu* (*dziad*/*dziadzio* 'grandfather'), *Władziu* (*Władzio*), *bobasu* (*bobas* 'infant'), *brudasu* (*brudas* 'slob') [besides the regular *bobasie*, *brudasie*], (f.) *matusiu* (*matuś* 'mother'), *Zochu* (*Zocha*), *Tolusiu* (*Toluś*); Ukr. (m.) *didúnju*/*didúsju* (*did* 'grandfather'), *báten'ku* (*bát'ko* 'father'), *tátu* (*tato* 'father'), (f.) *babúnju*/*babúsju* (*bábka* 'grandmother'), *mamúnju*, *Marúsju* (*máma* 'mother'). In Bulgarian, Macedonian and Serbo-Croatian the generalized ending is *-o* or *-e*: Bulg. (m.) *síne* or *sino* (*sin* 'son'), *číče* (*čičo* 'uncle'), *drugár'o* or *drugár'u* (*drugár* 'comrade'), (f.) *Maríjke* (*Marija*), *móme* (*móma* 'girl'); Maced. (m.) *tate* (*tatko* 'father'), (f.) *majke* (*majka* 'mother'), *tetke* (*tetka* 'aunt'); S-Cr. examples have been quoted above (6.2) Also numerous are forms with a zero ending; e.g. Russ. *mam* (*máma* 'mother'), *tët'* (*tëtja* 'aunt') *djad'* (*djádja* 'uncle'), *mátuš*(*k*) (*mátuška* 'mother'), *dev*(*k*) (*dévka* 'girl,

[20] The example is A. Obrębska's in "Technika spieszczeń w dzisiejszej polszczyźnie," *Język Polski*, XIV: 3 (1929), 65 ff.

wench'), *Van'* (*Ványa*), *Pet'* (*Pétja*), *Son'k* (*Són'ka*)[21] Polish examples have been listed (6.2).

8. The phenomenon of grammatical neutralization in expressive substantives allows us to draw some conclusions which are of a general linguistic, typological and historical relevance.

The obliteration of grammatical, functional and formal distinctions is a typical feature of expressive forms. It can be viewed as a special, extreme case of the phenomenon of neutralization, which serves to sharpen the polarity between various categories and classes. In addition to the prevailing grammatical structure of a language, it enables us to postulate a marked, "expressive grammar," the main feature of which is its grammatical under-differentiation compared to the more highly differentiated system of non-expressive forms. The various types and degrees of neutralization can be predicted and derived from that "broader" grammar by a set of rules. The rules vary according to the lexical-derivative structure of the expressive forms, which determines their emotive content.

The general phenomenon of "expressive grammar" does not obviate the necessity of describing in detail its implementation in various linguistic systems. The Slavic languages, with their complex "synthetic" structures and rich inventory of emotive devices, provide a very suitable framework for its discussion, but the phenomenon is by no means confined to Slavic. Other languages provide their own examples; cf. the gender of German diminutives *das Tischlein*, *Mütterlein*, *Häuschen*, from *der Tisch* ('table'), *die Mutter* ('mother'), *das Haus* ('house'), or of Modern Greek diminutives *τὸ σφαιρίδι*, *τὸ σκυλάκι* from *ἡ σφαῖρα* 'ball', *ὁ σκύλος* 'dog'.

The above exposition can be viewed not only as a statement of the various possibilities of grammatical neutralization in expressive substantives, but as a typological outline of the contemporary Slavic "expressive systems." Particularly revealing in this typology are the interdependence of grammatical neutralization and expressive types and the oscillations along language or dialect boundaries, as well as within one and the same literary system.

These oscillations and the existence of various types of neutralization reflect also the diachronic dimension of "expressive grammar." The grammatical development of Slavic expressive substantives was, as a matter of fact, a history of oscillation between various types of neutralization or of integration of the expressive "deviants" into the grammatical pattern of non-expressive forms. The history of the Russian expressive

[21] Cf. S. Obnorskij, "Die Form des Vokativs im Russischen," *Zeitschrift für slawische Philologie*, I (1925), 105 ff.

derivatives is a case in point. As is known, the Old Russian derivatives *domíšče* (*dom* 'house'), *ulicíšče* (*úlica* 'street'), *synísko* (*syn* 'son'), *dočeríško* (*doč*' 'daughter),' *Vánjuško* (*Iván*), *odežónko* (*odéžda* 'clothes'), *djád'ko* (*djadja* 'uncle'), *Demídko* (*Demid*), *ob'edálo* 'glutton', *gromílo* 'burglar', *zapravílo* 'boss', were originally of neuter gender (up to the XVI-XVII century), as they still are in various Russian dialects, especially the Northern.[22] The derivatives of type I neutralized the gender of the basic stem. Subsequently the gender of the basic stem became dominant in the derivative forms without, however, affecting the loss of the *-o* desinence; the result was a switch from functional neutralization to formal neutralization. The derivatives of type II, on the other hand, which shared the general Russian tendency to abolish animate neuter substantives, passed en masse to the *-a* declension, losing all but their animate gender in the plural (the 'epicoena'). The modern literary language rarely admits masculine substantives of this type, with the *-o* desinence in the nominative singular. Thus the two types have merely switched from one form of neutralization to another. On the other hand, some derivatives which were originally neuter (all forms with the expressive suffixes *-enko*, *-ko* and most forms with the suffix *-iško*) passed to the *-a* declension, adopting the gender of their base forms.

In the light of the above hypotheses, the various attempts to explain phonetically the development of the Russian expressive forms lose their validity (as has been suggested also by other scholars), while the older, piecemeal grammatical explanations are placed within a general linguistic and Slavic context.[23] It should, furthermore, be noticed that the history of "expressive grammar" closely resembles other diachronic processes in which certain relations remain constant in spite of their variant implementations, whereas other relations are eliminated or replaced by new relations.

Department of Linguistics
University of Chicago
Chicago 37, Illinois

[22] Cf. S. Obnorskij, *Imennoe sklonenie v sovremennom russkom jazyke*, I (1927), 18–23; B. O. Unbegaun, *La langue russe au XVI-e siècle* (1935), pp. 57 ff.; P. S. Kuznecov, *Istoričeskaja grammatika russkogo jazyka* (1943), p. 114.

[23] An historical survey of the various phonetic and morphological interpretations of the *-o*/*-e* endings in masculine proper names is found in Z. G. Rozova's "Kolebanija v sklonenii ličnyx imen mužskogo roda na -o i na -e v serbskoxorvatskom jazyke sravnitel'no s russkim," *Voprosy Russkogo Jazykoznanija, L'vovskij Gos. Univ. im. I. Franka*, III (1958), 132 ff.

JAN VOORHOEVE

Le ton et la grammaire dans le Saramaccan

1. Le but de cette étude.

Le Saramaccan est une langue créole que parlent les descendants des esclaves nègres ou marrons qui s'évadèrent autrefois; ils vivent maintenant sur les bords du Haut Surinam sur la côte Nord de l'Amérique du Sud. C'est la seule langue créole connue jusqu'à l'heure actuelle dans laquelle le ton est nettement distinctif.[1] Ceci est d'autant plus remarquable que la langue créole apparentée, le Sranan, que parlent dans la plaine, sur la côte de Surinam, les descendants des anciens esclaves nègres, n'est pas une langue tonale.[2]

J'ai déjà démontré dans l'analyse phonématique du Saramaccan[3] que le système compliqué qui comprend deux niveaux de ton, trois contours de ton et la durée des voyelles peut s'analyser comme un système de deux niveaux de tons (aigu et grave) relativement très simple, dans lequel chaque syllabe ou mot doit porter *un* niveau du ton.[4] Ceci n'est pas une erreur, mais une nette simplification des faits. On constate par exemple un

[1] Comparez *lǒ* 'clan' et *lò* 'mucosité', *dá* 'donner' et *da* 'être'. Voir pour l'orthographe note 6.

[2] Nous ne parlerons pas ici en détail de la différence entre une langue tonale (Saramaccan) et une langue à accent (Sranan). On a admis jusqu'à présent que cette différence est déterminée par la différence entre hauteur et accent. Des recherches récentes ont rendu vraisemblable qu'il n'y a pas de différence essentielle entre les deux et que l'accent est perçu lui-même par un auditeur via une différence de hauteur du ton. Voir à ce sujet, e.g. H. Mol et E. M. Uhlenbeck, 'The Linguistic Relevance of Intensity in Stress.' *Lingua* V (1955), pages 205–213. Ceci donne à notre problème une actualité particulière. Sans approfondir la question plus avant, je tiens pourtant à mentionner ici que la différence entre langue tonale et langue à accent (telle qu'elle se révèle dans la comparaison du Saramaccan avec le Sranan) repose sur l'organisation des différences de la hauteur du ton dans le mot. En Saramaccan, la syllabe est porteuse de la hauteur tonale (pour autant que cette dernière soit reconnaissable par sa relation avec d'autres syllabes), tandis qu'en Sranan c'est le mot qui est porteur d'une différence de hauteur tonale.

[3] J. Voorhoeve, 'An Orthography for Saramaccan.' *Word* XV (1959), pp. 436–445.

[4] Le contour du ton s'est révélé constamment parallèle avec la durée de la voyelle. Le comportement des voyelles longues a révélé que ces voyelles sont construites par deux (ou plus) voyelles brèves. Quand ces voyelles brèves possédaient des hauteurs de ton

niveau moyen lequel ne distingue à vrai dire aucune unité lexicale mais joue pourtant un rôle dans la syntaxe. On trouve ensuite un système de perturbations qui révèle la valence syntaxique de mots et de groupes de mots. On a donc affaire au-dessus du niveau des unités lexicales, en grammaire et en syntaxe, à des symptômes tonals, dont il n'a pas été question dans l'analyse des phonèmes.

Une étude spéciale qui éclairera ces facettes du comportement tonal des mots du Saramaccan permettra de faire remarquer la relation très étroite entre le ton et la structure grammaticale de la langue que je vais décrire. Nous étudierons successivement et plus amplement dans quelques paragraphes la relation entre ton et syntaxème, ton et structure du mot, ton et syntaxe.

2. TON ET SYNTAXÈME.

J'ai exposé déjà, en d'autres circonstances,[5] en rapport avec l'analyse grammaticale du Sranan, que la proposition ne peut être le point de départ de l'analyse grammaticale parce qu'elle n'est pas une unité grammaticale mais une unité d'intonation. J'ai alors défini le syntaxème comme la plus grande unité grammaticale comprenant une relation linguistique démontrable entre les parties composantes.

Dans le Saramaccan, le syntaxème est fermé par une influence descendante sur la dernière syllabe lors de syntaxème affirmatif, et sur les deux dernières syllabes lors de syntaxème négatif. Cette influence descendante agit selon des règles déterminées que quelques exemples vont tout d'abord rendre bien nettes.[6] Le cadre en usage du syntaxème affirmatif est:

différentes, il en résultait une voyelle longue avec contour tonal. En Yoruba, la même chose a été constatée par une autre voie. Voir B. Siertsema, 'Problems of Phonemic Interpretation, II. Long Vowels in a Tone I anguage.' *Lingua* VIII (1959), pages 42–64.

[5] Dans un article sur 'Linguistic Experiments in Syntactic Analysis,' lu à la Conference for Creole Language Studies à la Jamaïque (1959). Cet article sera publié dans les *Creole Language Studies*, 2e partie (à paraître).

[6] Le système orthographique employé est démontré ici. Pour de plus amples renseignements, voir l'article en question dans *Word*, XV.

I. Voyelles:

				i				u		
				e				o		
				è		ò				
				a						

Les voyelles nasales sont indiquées par voyelle plus *n* (avant des consonnes non-labiales) ou *m* (avant des consonnes labiales).

II. Consonnes:

p	b	kp	gb				m	mb	w	
				f	v					
t	d			s	z	n	nd			
tj	dj					nj	ndj			
k	g						ng	h		

a dú wán sondí ku . . . 'il a fait quelque chose avec . . .' Dans ce cadre, la syllabe *so(ndí)* reçoit un ton élevé sous l'influence du sandhi tonal dont il sera plus amplement parlé. La place vide est occupée par les mots et groupes de mots suivants:

forme de perturbation	*forme primitive*	*signification*
sábi	sábi	connaissance
mí	mí	moi
mí tata	mí *ta*tá	mon père
azanganá	azanganá	tibia
mí azangana	mí *azanga*ná	mon tibia
bò	bŏ	arc
dí bò	dí bŏ	l'arc
kŏnu bò	kŏ*nu* bŏ	arc du roi
so sòmbè bŏ	so sòmbè bŏ	l'arc (de) certaines personnes.

Le système des perturbations ne peut être interprété que si nous savons que les syllabes en italique dans la forme primitive ont subi une influence ascendante sous l'influence du sandhi tonal quand la syllabe finale n'a pas subi d'influence descendante par suite de sa position à la fin du syntaxème. Si nous intégrons les syllabes en italique dans le système, nous voyons que l'abaissement jusqu'au ton grave ne se présente que si aucune syllabe basse inaltérable n'a précédé. Ceci nous conduit à la règle suivante: *Quand un syntaxème affirmatif se termine sur une syllabe de ton aigu et qu'il est précédé par une syllabe de ton aigu ou par une syllabe ayant acquis le ton aigu en dehors de la fin du syntaxème* (à nommer plus tard: ton grave influençable), *la dernière syllabe change son ton aigu en un ton grave.*

Tout ton aigu situé à la fin du syntaxème est par essence un ton moyen. Ceci découle du fait qu'en Saramaccan tous les intervalles situés à la fin du syntaxème sont plus petits que ceux du début du syntaxème.[7]

Le cadre employé pour le syntaxème négatif: *á dú wán sondí ku.* . . . 'il n'a pas fait quelque chose avec . . .'

forme de perturbation	*forme primitive*	*signification*
mi	mí	moi
mí tata	mí tatá	mon père
azangana	azanganá	tibia

III. Ton: ´ (ton aigu), zéro (ton grave). Quand la consonne nasale forme une syllabe à part, le ton grave est indiqué par le signe ´ sur la consonne.

Les combinaisons de consonnes ne sont pas possibles. Chaque voyelle forme une syllabe à part. Dans cet article, les perturbations tonales ne sont pas transcrites sauf quand elles sont expressément mentionnées.

[7] Selon une communication du Dr. Jack Berry, il en est ainsi dans beaucoup de langues de l'Afrique occidentale. Le ton moyen peut être considéré comme un "allotone" du ton aigu, conditionné par la fin du syntaxème.

forme de perturbation	*forme primitive*	*signification*
dí bò	dí bǒ	l'arc
sabi	sábi	connaissance
deesi	deési	médecine
dí deesi	dí deési	la médecine
ahalakpakpa	ahalakpákpa	cafard
agó-futu	agó-fútu	talon
agó-fútu-wojo	agó-fútu-wójo	cheville
dí ádúmáwewe	dí *adu*máwewe	la puce d'animal
adumáwewe	adumáwewe	puce d'animal
azǒnjènjè	azǒnjěnjè	chou palme

La syllabe italique dans la forme primitive est haussée sous l'influence du sandhi tonal. Dans les syntaxèmes négatifs, l'influence descendante agit sur les deux dernières syllabes du dernier mot. Si la dernière syllabe a un ton aigu, elle change en un ton grave sans égard pour ce qui précède. Si la dernière syllabe a un ton grave, l'avant dernière syllabe élevée change en une syllabe grave. Si un mot finit sur deux syllabes graves, rien ne change. Ceci conduit à la règle suivante: *Un syntaxème négatif finit sur un mot dont la dernière syllabe est transformée en un ton grave, ou bien l'avant dernière syllabe, si la dernière a déjà ton grave.*[8]

Il s'ajoute ici un complément: l'influence ascendante quand le syntaxème n'est pas encore fermé. Cette influence ascendante n'est pas tenue à s'exercer à un endroit syntaxique déterminé dans le syntaxème et on la rencontre partout où celui qui parle introduit une pause sans fermer le syntaxème. On la voit certainement à la limite entre proposition subordonnée et proposition principale, mais aussi, par exemple, entre un complément et la partie du syntaxème qui suit ce dernier. Normalement, ce devrait être *a dí feéda mi gó a fóto* 'ce vendredi-là, je suis allé en ville'. Toutefois, si le parleur intercale une pause après *feéda*, il donne un ton aigu à la dernière syllabe de ce mot. Ceci indique notamment de façon bien nette que le syntaxème n'est pas encore fermé sur ce mot. Il était possible en effet que *a dí feéda* ait formé par lui-même un syntaxème.

Cette étude peut naturellement être poursuivie puisque nous connaissons maintenant le comportement tonal des mots à la fin du syntaxème. On peut déterminer expérimentalement de façon bien simple, sur la base du comportement tonal des mots, où se trouve la limite entre les syntaxèmes (les plus grandes unités syntaxiques dans l'emploi de la langue). On possède

[8] On verra dans le paragraphe suivant que les mots se terminant sur deux syllabes aigües appartiennent à une catégorie très particulière de mots d'intensification; ces mots sont caractérisés par le fait que la syllabe finale peut être allongée. A la fin d'un syntaxème affirmatif, un mot de ce genre ne subit aucun changement. A la fin d'un syntaxème négatif, la syllabe finale change en un ton grave et peut être aussi allongée en un ton grave.

ainsi, dans l'analyse du Saramaccan, une ressource pleine de valeur. Ce n'est point toutefois le but de cette étude de donner une analyse syntaxique du Saramaccan. Son but est atteint dès que la relation entre ton et syntaxème est déterminée.[9]

C'est avec intention que je n'ai point parlé ici d'un modèle d'intonation "super-imposed." L'intonation, en effet remplit encore un autre rôle que la caractérisation de la fin de la phrase (même si nous considérons grosso modo comme semblables le syntaxème et la phrase). L'intonation peut aussi distinguer les uns des autres des types de phrase déterminés. Dans une description structurelle du Sranan encore à publier, j'ai fait remarquer que, par le moyen de l'intonation, il est fait une distinction entre la proposition interrogative sans particule interrogative et la proposition communicative plus la proposition interrogative avec particule interrogative. J'ai observé aussi une même distinction dans le Saramaccan où elle n'est point toutefois caractérisée par une différence d'intonation mais par l'emploi de certains suffixes de syntaxème.

La proposition sans particule interrogative est caractérisée par un suffixe de syntaxème *-ŏ*[10], la proposition interrogative avec particule interrogative (ou bien avec un mot formé au moyen du préfixe interrogatif *ún-* qui devient ainsi particule interrogative) et la proposition communicative sont caractérisées par le suffixe de syntaxème zéro. Ceci est une règle impérieuse. On ne peut, par exemple, faire terminer la phrase interrogative avec particule interrogative sur le suffixe *-ŏ* et non plus négliger ce suffixe après la proposition interrogative sans particule interrogative. Finalement on trouve encore un suffixe de syntaxème *-jé* ou *-é* après une proposition exclamative, mais il peut éventuellement être négligé.

Ceci aussi nous montre clairement que les moyens d'organisation choisis par une langue tonale sont totalement autres que ceux choisis par une langue avec accentuation. On peut souvent encore, quand il s'agit de jeunes langues créoles, se rendre compte de la façon dont ces moyens d'organisation se sont établis. On peut ainsi démontrer que le suffixe *-ŏ* doit provenir de la particule interrogative *nŏ* 'n'est-ce pas?' connue aussi dans le Sranan à la fin de la phrase, et que le suffixe *-é* ou *-jé* doit provenir luit aussi de l'encouragement connu de même en Sranan *jére* ou *jéi* 'allons! (litt.: entends!)'.

[9] Le comportement tonal d'appositions, propositions incidentes, propositions contenant un discours direct ou un discours indirect, et des propositions subordonnées relatives, aura surtout de l'importance dans la poursuite de ces recherches.

[10] Ceci concorde de nouveau avec des variations de ton déterminées. Un syntaxème négatif suit la règle normale, à condition toutefois que le suffixe ne soit pas considéré comme syllabe finale. Dans un syntaxème positif, le sandhi tonal prend toujours place entre le suffixe et les syllabes précédentes.

3. Ton et structure du mot.

Le terme forme primitive dont nous avons déjà fait usage ci-dessus est resté indéfini. La forme primitive d'un mot est cette forme tonale du mot d'où peuvent dériver toutes les formes de perturbation. Le mot dont la forme primitive est *deési* 'médecine' peut présenter les formes de perturbation ci-dessous:

deesi (à la fin d'un syntaxème négatif),
déési (avec sandhi tonal au commencement du mot: *di déési*, 'la médecine',
deési (avec sandhi tonal à la fin du mot: *mi tá tjá deési dá di ómi*, 'je porte médecine pour l'homme',
déési (avec sandhi tonal au commencement et à la fin du mot: *mi tá tjá di déési dá di ómi*, 'je porte la médecine pour l'homme'.

Toutes ces formes perturbatives sont concevables vu la forme primitive du mot et les règles de perturbation. Toutefois la forme primitive n'est pas, sans plus, à déduire d'une forme de perturbation.

A côté du terme 'forme primitive', le terme 'forme isolée' est aussi important, c'est la forme de perturbation du mot, celle que possède le mot en position isolée (comme syntaxème positif par conséquent). Cela veut donc dire que la syllabe finale de ton aigu change en ton grave quand la syllabe précédente a un ton aigu ou un ton grave influençable. Le redoublement *gasá-gasá* 'très vieux' a une forme isolée *gasá-gasa*, parce que l'avant dernière syllabe a un ton grave influençable; le redoublement *njú-njú* 'nouveau' a une forme isolée *njú-nju*, parce que l'avant dernière syllabe a un ton aigu. Le terme 'forme isolée, est très important quant aux mots composés et aux redoublements. Nous remarquons dans les 'one-morph-words' que la forme primitive coïncide généralement avec la forme isolée. Il y a toutefois d'importantes exceptions, notamment dans les mots se terminant par deux (ou plus) syllabes élevées. Nous avons trouvé dans le Saramaccan les possibilités théoriques suivantes (s = syllabe, ´ = ton aigu, zéro = ton grave):

śś *fáán* et autres mots destinés à intensifier, plus *déé* (article du pluriel)
sśś *abíí* ' que désirez-vous?'
śśś *gbólóó* et autres mots destinés à intensifier.
ssśś —
sśśś —
śśśś —
śśśś *kólólóó* et autres mots destinés à intensifier.

Aucune des possibilités théoriques offertes par les mots de cinq et six syllabes n'a été employée. Je n'ai trouvé aucun 'one-morph-word' ayant plus de six syllabes. Je dois faire remarquer ici que, par suite de l'extension de la dernière syllabe dans les mots d'intensification, il est fort possible que

la dernière syllabe ne doit compter que pour *une* syllabe.[11] Je n'ai pas poursuivi mes recherches en ce sens ayant bien vite constaté que ces mots d'intensification pouvaient être comparés dans leur comportement tonal aux mots d'*une* syllabe. L'ensemble du matériel étudié comprend environ 1500 mots.[12] Nous avons tenu en dehors de ces recherches les mots composés, redoublements, et dérivés pour des raisons dont nous parlerons plus tard.

L'aperçu acquis révèle nettement que les structures tonales se terminant sur deux, ou plus, de syllabes aigües (structures de ton dans lesquelles la forme primitive dévierait de la forme isolée) restent soit inutilisées, soit réservées à des mots d'intensification, lesquels sont sans signification lexicale spéciale et destinés seulement à cette intensification. C'est ainsi que *fáán* indique l'intensité de 'blanc', *gbólóó* indique l'intensité d'une assertion précédente, *kólólóó* renforce l'intensité du départ (départ définitif). Il y a aussi différents mots pour exprimer l'intensité de 'regarder': fixer avec attention, surveiller étroitement, regarder avec étonnement etc. Ces mots ne possèdent pas de signification lexicale propre et ne peuvent pas se présenter isolés. Ils ont tous, en même temps, cette propriété que leur dernière syllabe peut recevoir de l'extension de sorte que l'on n'est jamais sûr du nombre des syllabes qui composent le mot. Le mot *abíí* possède aussi cette propriété; c'est le seul mot construit avec *sśś*. Il ne peut être compté parmi les mots d'intensification, ne s'emploie que comme réaction à une interpellation et seulement isolé. La seconde exception est formée par l'article pluriel *déé*, lequel est 'phrasally bound' et ne peut pas être isolé. Il n'y a donc point, dans tous ces cas, possibilité de trouble entre la forme primitive et la forme isolée. J'en viens donc à conclure ici que les différentes possibilités sont ainsi partagées que, dans le Saramaccan, la forme isolée ne peut jamais dévier de la forme primitive, du moins si nous nous en tenons aux 'one-morph-words'.

Les informateurs donnent toujours la forme isolée des composés, redoublements, etc., laquelle ne doit donc pas nécessairement concorder

[11] Nous avons signalé ceci à propos du mot *gbólóó*. A la fin d'un syntaxème négatif on trouve: *á dĕ sŏ gbóloo* 'ce n'est sûrement pas ainsi'. Ceci semble donc être un mot de deux syllabes avec extension de la syllabe finale. Je n'ai pas jugé nécessaire de contrôler ce fait quant à tous les mots d'intensification, ce qui est d'ailleurs difficile à faire, parce que beaucoup d'entre eux ne peuvent être employés dans un syntaxème négatif. Il est suffisant, quant au but que je me propose, de décrire la structure tonale déviant de ce type de mots.

[12] Ce matériel de mots a été réuni pour la plus grande partie par le Père A. Donicie, missionnaire à Suriname. Le matériel sera publié à part avec une étude lexicographique comparative: comparaison du Saramaccan de 1778 avec celui de l'heure actuelle et comparaison du Saramaccan avec le Sranan.

avec la forme primitive. Cette dernière doit toujours dériver des diverses formes de perturbation, ce qui peut amener de grandes complications parce que tous les mots ne sont pas immédiatement reconnus comme des composés. Les composés les plus faciles à juger sont ceux dont est connu un élément hors du composé: *nángo* 'aller' et *gaánduwe* 'il y a longtemps'. Le mot indiqué comme *nángo* possède selon le comportement tonal la forme primitive *nángó*. Le mot concorde entièrement avec le groupe de mots *tá gó* (le verbe *aller*, précédé de la particule verbale *tá* comme indication de l'aspect duratif du verbe). Nous admettons que *nán* soit une indication vieillie de l'aspect duratif laquelle n'a subsisté que dans ce seul composé. Le mot cité sous la forme de *gaánduwe* possède l'élément *gaán* 'grand'. Une forme primitive *gaánduwé* peut dériver des formes de perturbation et je suppose par conséquent qu'il y a ici un vieux composé: *gaán-duwé*, dont *un* élément ne se présente plus hors du composé actuel.

Il est plus difficile de décider quand on ne peut reconnaître, dans le mot, aucun élément composant. J'ai trouvé trois mots, indiqués *kúnunu* 'montagne', *bódjee* 'rusé' et *kúndjee* 'sorte d'arbre'. On peut déduire des formes de perturbation une forme primitive: *kúnunú*, *bódjeé* et *kúndjeé*. On peut éventuellement découvrir encore ici des éléments communs: *kún-* et *-djeé* notamment. Un élément *kún-* se terminant sur une consonne est très invraisemblable en Saramaccan. Il est toutefois possible de supposer un élément *kúnu-* dérivé en *kún-* dans le dernier composé. Sur la base de ces considérations, je me les représente comme d'anciens composés sous les formes suivantes: *kúnu-nú*, *bó-djeé* et *kún(u)-djeé*.

Je me suis basé dans ces considérations sur ce principe que la structure tonale des 'one-morph-words' du Saramaccan ne permet point l'emploi de syllabes basses entre des syllabes aigües. Lors d'une liaison syntaxique intime à la limite de deux mots, des syllabes graves placées entre des syllabes aigües changent en syllabes aigües sous l'influence du sandhi tonal: *dí alísi* 'le riz' devient sous cette influence *dí álísi*. On observe aussi ce même symptôme à l'intérieur du composé. Un composé d'*alísi* 'riz' et *páu* 'tige' devient *alísí-páu* 'tige de riz'. Je suis parti de la supposition qu'il n'y a pas de liaison plus solide que celle établie à l'intérieur d'un 'one-morph-word' et que, là, il n'est sûrement pas possible qu'une syllabe grave s'introduise entre des syllabes aigües.

Cette supposition n'est pas complètement confirmée par le matériel de mots qui a été étudié. J'ai publié ici toutes les exceptions constatées et donne pour chaque mot deux formules de structure (tout d'abord la forme indiquée par les informateurs et ensuite la forme de perturbation supposée quand ce mot s'est présenté isolé; par ailleurs, les règles du sandhi tonal sont appliquées à l'intérieur du mot aussi). La forme de perturbation

concorde avec la forme primitive d'un autre mot, de sorte qu'une méprise est possible entre la forme primitive et la forme de perturbation .

asóboní 'nom tabou de la vache'	sśsś–śśss
obílogbén 'sorte de serpe'	
kakáabá 'nom tabou du chien'	
gbégbédesú 'nom tabou du cerf'	śśsś–śśss
kpókpóosú 'avec les jambes croisées'	
asakúlubú 'nom d'une sorte de riz'	ssśsś–ssśss
asióboní 'variante de *asóboní*'	
adjúabèè 'nom tabou du tigre'	sśssś–śśsss
adjáintóo 'nom d'une sorte de riz'	sśsśs–śśśśs
obílogbéin 'variante de *obílogbén*'	
akúkulukumbá 'nom tabou du tigre'	sśssśś–śśssss
ahútúkúmabú 'surnom du tapir'	sśśśśś–śśśśss
ahútukúmabú 'variante de *ahútúkúmabú*'	sśśśśś–śśśśss

Le nombre de 10 mots sur un matériel de 1500 est sans aucun doute fort peu de chose. De plus, il est frappant de constater que, sur ces 10 exceptions, on compte 5 noms tabous, 1 surnom et 2 noms de sortes de riz indigène. Les noms tabous et les surnoms sont rarement employés dans une phrase et leur comportement tonal est difficile à contrôler. Le mot *obílogbén* peut avoit été à l'origine un nom tabou ou un surnom. J'ai rencontré plus tard le seul mot un peu moins exceptionnel *kpókpóosú* sous une forme complètement semblable à la forme de perturbation reconstruite ici, notamment *kpókpóosu*. Il n'est pas invraisemblable que ce soient des restes d'une langue africaine ayant d'autres structures tonales, lesquels ont pu conserver leur ancienne structure tonale à cause justement de leur présentation isolée (en tant que noms). Ces exceptions ne sont donc point pour moi des raisons de modifier ma supposition que le Saramaccan n'autorise pas de structures de mots avec syllabes graves placées entre des syllabes aigües. Ce qui veut dire par conséquent que, dans les 'one-morph-words,' la forme primitive coïncide avec la forme isolée.

Les dérivés se terminant par *-ma* et *-wan* ont en commun cette particularité qu'ils ne sont pas soumis à l'influence du sandhi tonal, mais bien aux perturbations qui marquent la fin ou la non-fin du syntaxème.[13] Ces suffixes ont un ton grave après un mot primaire possédant une syllabe

[13] Cette formation du mot est productive. De nouvelles formations comme *Faliséma* 'Pharisien' se révèlent suivre la règle. On peut dire, en général, que les dérivés de *-ma* renferment les nomina actoris et sont par conséquent dérivés de verbes, et que les dérivés de *-wan* sont des adjectifs pris substantivement. J'écris ceci avec une certaine hésitation parce que, partant de la forme du mot seulement, il est difficile de faire une distinction entre verbes et adjectifs. Peut-être devrions-nous parler de verbes d'action et de verbes exprimant une situation.

avec ton aigu et un ton aigu après un mot primaire n'ayant pas de syllabes à ton aigu.

zéngěnima 'hamac', du mot primaire *zèngěni*, 'planer',
heépima 'aide, sauveur', du mot primaire *heépi*, 'aider',
kondáma 'témoin', du mot primaire *kondá*, 'raconter',
ótowan 'autre', du mot primaire *óto*, 'autre',
búnuwan 'bon', du mot primaire *búnu*, 'bon',
donúwan 'jaune', du mot primaire *donú*, 'jaune',
lègèdèmá 'menteur', du mot primaire *lègèdè*, 'mentir',
takuwán 'mauvais', du mot primaire *taku*, 'mauvais',
bòsò-bòsò-wán 'ce qui n'est pas fortement lié', de *bòsò-bòsò*, 'détaché'.[14]

Les redoublements se comportent pour la plus grande partie comme des composés, ce qui veut dire que l'on double le mot singulier dans sa forme primitive et que ce redoublement est soumis à toutes les règles de perturbation. Le mot *avó* 'grand-parent' a le redoublement *avó-avó* 'différents ancêtres', le mot *hía* 'beaucoup' redouble en *hía-hía* 'beaucoup-beaucoup', le mot *wánté* (qui est lui-même un composé signifiant 'tout de suite') redouble en *wánté-wánté* 'immédiatement', le mot *gadja* 'long, majestueux' redouble en *gadja-gadja* 'avec grande puissance', le mot *baafu* 'soupe' redouble en *baafu-baafu* 'comme de la soupe', le mot *baáka* 'noir' redouble en *baáka-baáka* 'bien noir', le mot *baziá* 'descendre' redouble en *baziá-baziá* 'prosterné' etc.[15] En ce qui concerne certaines formes de redoublement telles que *gasá-gasá* très vieux on n'en connaît pas la forme singulière, toutefois ces mots suivent les règles générales.[16] On trouve parfois de petites déviations dans la formation du redoublement. C'est ainsi que *awoo* vieux redouble sous la forme *awoo-woo* 'vieillot' et le mot *búnu* 'bon' donne la forme *búm-búu*, à côté de *búnu-búnu* 'très bon', régulièrement formé. Toutefois ceci n'atteint nullement la structure tonale de la forme primitive et nous ne nous arrêterons plus à ces exceptions.

[14] Le mot *gaamá* 'gouverneur, chef' est toutefois soumis aux règles du sandhi tonal et, pour le reste, il est construit de façon divergente (mot primaire: *gaán*, 'grand'). Ce mot doit être emprunté au Sranan *gramán*. Construit de même de façon divergente, citons le mot *íniwán* 'chaque, chacun et tous' qui a deux syllabes aigües et une intermédiaire grave.

[15] Les redoublements sont ici, eux-aussi, transcrits sous leur forme primitive de laquelle toutes les formes de perturbation peuvent dériver. Les informateurs donnent toutefois la forme divergente isolée, tandis qu'en même temps les syllabes graves entre les syllabes aigües sont changées en syllabes aigües: *avóavo*, *híáhía*, *wántéwánte*, *gadja-gadja*, *baafubaafu*, *baákábááka* et *baziábazia*.

[16] On indique *gaságasa*. Voir aussi la forme *avóavo*. Dans cette forme isolée, de même qu'à la fin du syntaxème, la syllabe finale à ton aigu est changée en ton grave, parce que la syllabe grave précédente est placée entre deux syllabes aigües, et qu'elle a donc un ton grave influençable.

Il arrive aussi que le redoublement possède une forme primitive qui n'a pu se produire par le redoublement de la forme primitive du mot singulier. La partie du matériel (env. 900 mots) qui a été étudié a fourni les exceptions suivantes (toujours données sous la forme isolée): *biibíi* 'plaine', *djombodjómbo* 'grenouille', de *djómbo*, 'sauter', *gobogóbo* 'sorte d'arachide', *sakasáka* 'balle', *hènèhènè* 'miettes', *hònjòhŏnjò* 'mouche de vache', *gbonogbóno* 'sorte de mousse', *kasikaási* 'ulcère, blessure', de *kaási*, 'démangeaison', *kesikési* 'sorte de singe' *kòŏkŏò* 'tousser', *kĕdĕkèdè* 'juste, précis, prudent', *lĕbĕlèbè* 'long et maigre', *kĕlĕkèlè* 'tout à fait, *vĕnĕvènè* 'nu comme un ver'.

Si l'on pense qu'en principe tous les verbes, tous les adjectifs, et une partie des substantifs peuvent former des redoublements toujours formés régulièrement, ceci n'est qu'un petit nombre d'exceptions, Il est de même frappant que seuls deux redoublements puissent être mis en rapport avec un mot singulier.

Les premiers redoublements sont ainsi formés qu'*une* syllabe seule a un ton aigu. La forme isolée coïncide ici en même temps avec la forme primitive ce qui n'est le cas, dans les redoublements régulièrement formés, que lorsque le mot singulier ne contient pas de ton aigu. Ces redoublements suivent complètement les principes de formation du mot dans le Sranan, langue à accent avec accent sur le mot. Les mots suivants ont un équivalent Sranan de signification concordante; (l'accent du mot est symbolisé de même dans le Sranan, par le signe ′): *biibíi* 'biribíri', *sakasáka* 'sakasáka', *kasikaási* 'krasi-krási', *kesikési* 'kesikési'. Etant donnée cette analogie, il me parait justifié de supposer que ce groupe de redoublements de construction différente est emprunté au Sranan et, par suite, présente aussi la construction des redoublements du Sranan.

Le mot *kòŏkŏò* semble être, dans son cours tonal, une imitation de son et nous ne nous en occuperons pas plus avant. Les quatre derniers redoublements ont en commun cette particularité que leur forme primitive ne contient que des syllabes aigües. A l'encontre de ce qu'on pourrait attendre, ce n'est pas la syllabe finale qui est abaissée dans la forme isolée mais tout le second élément du redoublement. Ces redoublements sont tous placés à côté d'un autre mot non redoublé, soit respectivement *kĕdĕĕ*, *lĕbĕĕ*, *kĕlĕlĕĕ* et *vĕnĕnĕĕ*, lesquels appartiennent, pour la forme et pour la signification, au groupe des mots d'intensification ainsi nommés. Nous avons déjà signalé, lors de la discussion de ce groupe, que le nombre des syllabes n'a pu être déterminé exactement qu'avec difficulté. Dans la forme redoublée, la forme singulière se révèle traitée comme *une* seule syllabe. Ces anomalies correspondent par conséquent à d'autres anomalies de même genre dans la structure des mots du Saramaccan.

4. TON ET SYNTAXE.

Le sandhi tonal du Saramaccan n'est pas exclusivement conditionné par l'entourage tonal. Le sandhi tonal exerce une influence ascendante sur des syllabes graves à la limite du mot entre des syllabes aigües, mais cela dans des constructions syntaxiques spéciales seulement. Le mot *buúu* 'sang' a dans la construction *dí búúu* 'le sang' une autre forme tonale que dans la construction *dá buúu* 'donne du sang'. On ne trouve notamment aucun sandhi tonal entre le verbe et l'objet non-pronominal placé en arrière, mais on en trouve entre le substantif et le mot déterminatif placé immédiatement devant le substantif.

Le sandhi tonal rend évidente la valence syntaxique qu'ont les mots dans le syntaxème. Quand après la forme *dí* il vient un substantif et que l'on ne trouve pas un sandhi tonal entre les deux mots, on sait avec certitude que *dí* n'est pas un article mais une conjonction. Les symptômes du sandhi tonal donnent une aide précieuse lors de l'analyse syntaxique.[17] Ceci est illustré par la traduction d'Acta 21,15 en Saramaccan, dont tous les mots sont d'abord écrits dans leur forme primitive, et transcrits ensuite avec les perturbations:

ma dí déé dáka dě pasá hěn, u téi seeká, / u gó a Jelúsalèn.
ma dí déé dáká dě pásá hěn, u téi seeká, / u gó a Jelúsalèn.

Les perturbations révèlent que:

a. *déé dáka dě* est un groupe de substantif dont *dáka* est le centre,
b. *déé dáka dě* est le sujet du verbe suivant,
c. *seeká* est un objet de *téi* (et donc un substantif).

Je m'étendrai un peu plus sur ce fait. Si *seeká* était un verbe, il serait perturbé et devenu *seeka*. La syllabe finale ne reste élevée à la fin du syntaxème (symbolisé ici par /) que lorsque la syllabe précédente possède un ton grave non-influençable, c'est-à-dire quand aucun sandhi tonal ne se présente entre *téi* et *seeká*. Ceci n'est possible que lorsque *seeká* est objet du verbe précédent.

Le but envisagé par cette étude est de démontrer la relation entre le ton et la syntaxe et non point de donner une analyse syntaxique complète. C'est pour cela que nous décrivons seulement les symptômes du sandhi tonal dans les principaux groupes syntaxiques. Par ailleurs, le sandhi tonal n'est décrit en détail que dans le groupe du substantif (lequel se prête le mieux à cette étude). Nous faisons usage dans ces descriptions de

[17] L'analyse syntaxique sur la base des symptômes de sandhi tonal est plus simple que si cette analyse devait se faire sur la base de la possibilité d'un placement séparé des mots à l'intérieur du groupe de mots. Voir quant à l'explication de cette dernière méthode ma publication déjà citée *Linguistic Experiments in Syntactic Analysis*.

formules de sandhi tonal dans lesquelles les différentes catégories sont indiquées par des termes conventionnels ou par des exemples, et les symptômes de sandhi tonal par des symboles: + = sandhi tonal, − = pas de sandhi tonal, × = impossibilité de sandhi tonal par suite d'entourage tonal, ± = sandhi tonal avec perte de syllable.

I. *Le groupe du substantif:* (*híi* − article − numéral + *óto* + adjectif + détermination de nationalité) + substantif + *seépi* + *dě*/*sŏ* − *túu*/*akí*. Cette formule signifie que le sandhi tonal agit entre le substantif et chacune de ses déterminations ayant immédiatement précédé, quelle que soit la catégorie à laquelle appartient la détermination impliquée Ceci est symbolisé par les parenthèses suivies de +. Elle veut dire ensuite que, lors de plusieurs determinations précédentes, les symptômes de sandhi tonal que l'on trouve parmi ces expressions sont réglées comme l'indique la formule entre parenthèses. Enfin, la formule signifie encore qu'entre le substantif et la (ou les) détermination placée après lui, les symptômes de sandhi tonal sont réglés selon la formule. Les exemples suivants peuvent illustrer ce qui précède:

híi dí buúu > híi (−) dí (+) búúu 'tout le sang'.
dí potigé mujěè > dí (−) potigé (+) mújěè 'la femme portugaise.'
dí óto potigé mujěè > dí (−) ótó (+) pótígé (+) mújěè 'l'autre femme portugaise'.
déé mujěè akí > déé (+) mújěè (−) akí 'les femmes ici, ces femmes'.
déé mujěè dě túu / déé (+) mújěě (+) dě (−) túu 'les femmes là, toutes'.

On peut aussi employer cette formule lors de la détermination des limites du mot, surtout quand il s'agit de différencier entre un groupe de mots et un mot composé. Le groupe de mots *dí gaán wósu* est construit selon la formule dí (−) gaán (+) wósu 'la grande maison'. Quant au mot composé *gaán-wósu* 'maison avec chambre à coucher', la formule devient dí (+) gáán-wósu. Quand la première syllabe du composé a un ton grave, cette méthode est simple à appliquer.

La formule de sandhi tonal donnée ici quant au groupe du substantif doit être plus amplement développée. On peut, à la place de l'article, introduire aussi un substantif déterminatif ou un pronom personnel indiquant le possesseur du substantif:

dí kŏnu mujěè > dí kŏnú (+) mújěè 'la femme du roi'.
mí potigé mujěè > mí (−) potigé (+) mújěè 'ma femme portugaise'.

Il faut encore remarquer que l'on peut employer plusieurs adjectifs au

lieu d'*un* (voir le terme dans la formule). Dans ce cas on ne trouve pas de sandhi tonal entre les adjectifs:

dí gaán boóko wósu > dí (−) gaán (−) boókó (+) wósu 'la grande maison démolie'.

Finalement, la détermination indiquant le possesseur peut être placée aussi après le substantif (à la place de *seépí*), mais alors précédée par *u* ou *fu* 'de'. Devant un substantif, ce petit mot *u/fu* est considéré comme indépendant. Devant un pronom, il se comporte toutefois complètement comme un préfixe:

dí mujěè u mí > dí (+) mújěě (+) ú mí . . . 'la femme à moi . . .'.[18]

A la fin du syntaxème, on trouve cependant la forme de perturbation suivante: dí (+) mújěè (+) u mi. Ces formes de perturbation ne peuvent s'expliquer que si *u* est considéré comme une partie du mot *u-mí*. Dans le premier exemple, il ne pourrait y avoir aucun sandhi tonal à la limite entre les deux mots *mujěè* et *u*, et le mot *u* tout entier ne pourrait certainement pas avoir le ton aigu. Dans le second exemple, *mí* seul pourrait avoir un ton aigu si *ú* possédait un ton grave influençable et, par conséquent, formait une syllabe du dérivé *u-mí*.[19]

II. *Le groupe adjectif:* détermination − adjectif.

dí mŏò lóngi pási > dí (−) mŏò (−) lóngí (+) pási 'le plus long chemin'.

III. *Le groupe adverbial:* détermination + adverbe. Ou: adverbe + détermination.

mi wáka mŏò hési > mi wáka mŏŏ (+) hési 'je marchais plus vite'.

mi wáka hési mŏò í > mi wáka hésí (+) mŏò (−) í 'je marchais plus vite que toi'.

IV. *Le groupe déterminatif:* . *a* × objet de la détermination + spécification. L'objet de la détermination peut être un groupe de substantif ou un pronom. Il est très possible que la spécification même soit à considérer comme un substantif, de sorte que la formule serait réduite à: . *a* × groupe

[18] Les points signifient que le syntaxème doit être poursuivi. Dans l'exemple suivant, le point marque la fin du syntaxème.

[19] Ceci est le seul cas observé jusqu'à l'heure actuelle dans lequel on ait pu décider sur la base des symptômes de sandhi tonal qu'une syllabe est un préfixe. J'hésite encore sous ce rapport, parce que la forme *u* devant un substantif (dans la même signification exactement) et devant un verbe (dans une signification quelque peu différente) n'est pas un préfixe. J'ai pensé au début que la particule verbale *sa* était elle-aussi un préfixe parce que cette forme prenait un ton aigu après une négation. Une analyse plus approfondie a toutefois démontré une différence entre *sa* (désignation "irrealis") et *sá* 'pouvoir'. La première forme n'était pas possible après une négation, à moins qu'elle ne soit précédée de *bi* (désignation d'un temps passé).

substantif. Toutefois la spécification est alors un substantif de nature particulière, indiquant des directions déterminées à l'endroit du parleur: *líba* 'en haut, sommet', *déndu* 'entre', *bandja* 'à côté, côté', *báka* 'derrière, dos'.

a dí táfa líba > a (×) dí táfá (+) líba 'sur la table'.
a dí táfa bandja / a (×) dí táfa (+) bandja 'à côté de la table'.
a mí báka / a (×) mí (+) báka 'derrière moi'.
a déé sòmbè déndu / a (×) déé sòmbè (+) déndu 'entre les gens'.

V. *Le groupe acteur-action:* . (sujet − négation) + verbe 1 + verbe 2 − objet/détermination adverbiale + verbe 3. Cette formule n'est en vigueur que pour un sujet et objet non pronominal, ou pour un sujet et objet consistant en un pronom avec détermination plus précise. Quand le sujet est un pronom employé non emphatiquement sans détermination plus précise, les formules suivantes sont en vigueur: a) . sujet × verbe 1 . Ou: b) . sujet ± négation + verbe 1. Pour un objet pronominal, la formule suivante est en vigeur: . verbe 1 + objet + verbe 2. La 3ème pers. pl. a un ton grave comme objet, c'est pourquoi, dans cette formule + doit être remplacé par ×. Nous l'expliquerons plus à fond dans le paragraphe suivant.

Voici deux exemples pour illustrer ce qui précède:

dí ómi á ó toosá tatái dá dí miíi > dí ómi (−) á (+) ó (+) tóósá (−) tatái (+) dá dí miii 'l'homme ne va pas enrouler une corde pour l'enfant'.[20]
dí ómi ó toosá tatái > dí ómí (+) ó (+) tóósá (−) tatái 'l'homme va enrouler une corde'.

De nombreuses variations encore à la formule citée sont naturellement possibles, par exemple celle de donner une autre place à l'objet par rapport au verbe. Je ne puis toutefois décrire ici la syntaxe Saramaccane. Une analyse complète donnera sans aucun doute plus de lumière. Je n'ai pas encore pu, par exemple, résoudre le problème de la différence entre *mí seépi* 'moi-même, moi aussi' avec et sans sandhi tonal entre les deux éléments. A mon avis, j'ai pourtant réussi à démontrer dans ce paragraphe qu'il existe une relation manifeste entre le ton et la syntaxe dans le Saramaccan.

5. Le comportement tonal du pronom personnel:

Les pronoms personnels sont caractérisés par le fait qu'ils peuvent posséder un ton grave ou un ton aigu.[21]

[20] *dá* 'donner' est un verbe que l'on place devant ce qui serait un objet indirect dans le système linguistique européen. La forme *miii* est expliquée par les règles de perturbation à la fin d'un syntaxème négatif.

[21] Après les pronoms personnels, je crois n'avoir observé ceci que dans *kaa* 'déjà',

Les formes les plus usitées sont:

	sing.	*pl.*
1.	mi/mí	u/ú
2.	i/í	un/ún
3.	a/hěn	de/dé

On trouve aussi la forme *mí* comme *m̀*: *a m̀ báka* 'derrière moi'. Lorsqu'il s'agit d'un sujet devant une négation, il se présente une perte de syllabe: *mi á* > *m'ě* ou *m'á*, *i á* > *j'á*, *a á* > *á*, *u á* > *w'á*, *un á* > *w'ǎn*.[22] Le pronom ne forme plus une syllabe à part et n'a donc plus un ton propre. La forme *hěn* se voit aussi pourtant comme *ěn*, *hě* ou *ě*. La forme *ún* se voit aussi comme *únu*. La forme *í* possède de plus une forme *jú* (employé surtout avec des mots injurieux).

Nous avons déjà décrit dans le paragraphe précédent que les règles du sandhi tonal dans le groupe acteur-action sont différentes quand un sujet ou objet pronominal sans détermination plus précise se trouve à l'intérieur du groupe. En ce qui concerne les formules du sandhi tonal, le lecteur peut s'en référer au paragraphe précédent. Voici quelques exemples.

On trouve un sandhi tonal entre le verbe et l'objet pronominal suivant: *mi púu hěn kumútu* > mi púú (+) hěn (+) kúmútu 'je l'en ai retiré'. Ceci n'est pas le cas pour un groupe pronominal: *a púu hěn seépi kumútu* > a púu (−) hěn séépí (+) kúmútu 'il s'en est sorti'.

Le pronom employé non emphatiquement a toujours pour sujet un ton grave; en tant que détermination d'un substantif (indiquant le possesseur) toujours un ton aigu; comme objet déterminatif, toujours un ton aigu et comme objet un ton aigu, excepté à la 3ème personne du pluriel.[23]

mi gó 'j'allais'	*a púu mí kumútu* 'il m'en a retiré'
de gó 'ils allaient'	*a púu de kumútu* 'il les en a retirés'
mí wósu 'ma maison'	*a mí báka* 'derrière moi'
dé wósu 'leur maison'	*a dé báka* 'derrière eux'

Il y a diverses possibilités lors de l'emploi emphatique. On peut, par exemple ajouter *seépi* 'même' après les pronoms à ton aigu, ou *u* plus pronom plus *séepi*: *mí u mí seépi* 'moi pour moi-même'.

En position de sujet, il y a aussi possibilité d'une forme emphatique, laquelle semble provenir d'un redoublement du pronom, mais toutefois la

qui se présentait aussi comme *káá*. Il est toutefois possible que la forme primitive soit *kaá*. Ceci est difficile à trancher parce que ce mot se présente presque toujours à la fin du syntaxème.

22 *de á* n'est pas tout à fait contracté en *d'á*, mais la syllabe *de* est devenu extrêmement courte.

23 Dans ces exemples, les symptômes de sandhi tonal ne sont plus transcrits. Ils n'ont plus d'intérêt ici.

forme à ton aigu est toujours suivie de la forme à ton grave. Mais ces formes de redoublement peuvent subir aussi des modifications caractérisantes; c'est pour cela que nous donnons toutes les formes en schéma.

	sg.	*pl.*
1	mí-i	ú-u
2	í-i	ú-un
3	hĕn-a	dé-e

Le comportement tonal du pronom personnel dévie tellement tant en la forme (avec ton aigu ou ton bas en différentes positions syntaxiques) que dans les symptômes de sandhi tonal que cette catégorie grammaticale en est nettement caractérisée.

6. TON ET REPRÉSENTATION ORTHOGRAPHIQUE.

L'analyse phonétique révèle qu'en Saramaccan chaque syllabe consiste en *une* voyelle accompagnée ou non d'*une* consonne qui précède. Au niveau lexical, il y a ici peu d'exceptions: *adjàǹtóo* et *ńtóóbi*. Ces exceptions sont vraisemblablement issues de l'élision de la voyelle i: *adjámitóo* et *híntóóbi* (cette dernière variante a été rencontrée aussi dans le matériel des mots). Ceci concorde avec les résultats de l'élision de la voyelle *i* après une consonne nasale dans des groupes de mots: *duumí búnu* 'dors bien' › *duuḿ búnu*, *a mí báka* 'derrière moi' › *a ḿ báka*. Ce sont des exceptions explicables et strictement limitées. Nous avons de même déjà décrit la perte de syllabe dans la combinaison du pronom personnel et de la négation: *i á* › *j'á*. Ce sont ici aussi des exceptions strictement limitées. On peut donc établir en général que chaque syllabe a, au niveau lexical, son ton individuel.

Au niveau syntaxique, nous trouvons d'innombrables modifications tonales. Pike[24] conseille de ne pas écrire des variations insignifiantes. "On the other hand, tonemes substituted in morphology or syntax or sandhi should be written as pronounced". A mon avis, les raisons suivantes font que ceci ne saurait être appliqué au Saramaccan:

(1) la forme tonale primitive du mot n'est pas dérivable de la forme perturbée, tandis que la forme perturbé peut bien être réduite de la forme primitive; cela, étant donné les règles de perturbation.

(2) beaucoup de mots nettement à différencier entre eux dans la forme primitive (p. ex. *kái* 'appeler' et *kaí* 'tomber') ne peuvent plus l'être dans

[24] Kenneth L. Pike, *Tone Languages. A Technique for Determining the Number and Type of Pitch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion*. Ann Arbor, ₄1957, page 37. Dans une lettre personnelle, Pike appelle mon attention sur un récent article de Sarah Gudschinsky dans *Word*; il y est formulé une autre opinion qui, à son avis pourrait bien être une amélioration.

la forme perturbée (p. ex. *káí*). La forme primitive est plus nettement distinctive que la forme perturbative.

(3) ceux qui emploient la langue se rendent clairement compte de la forme primitive des "one-morph-words" (quand elle coincide avec la forme isolée), mais non point des formes de perturbation dont ils font usage.

(4) en gardant l'unité de l'image du mot dans toutes les situations (en ce qui concerne les signes tonals aussi) on suit dans les langues tonales le même système d'orthographe que dans les langues à accent, dans lesquelles, notamment, les symptômes de sandhi à la limite des mots ne sont pas rendus visibles dans l'orthographe.

J'ai pleine conscience de ce que le système orthographique que j'ai suivi est très compliqué pour des étrangers devant apprendre à manier automatiquement les règles de perturbation décrites ci-dessus. Sans la connaissance de ces règles, on ne peut lire une seule phrase sans faute en Saramaccan. Quand j'ai eu passé des mois à traduire du Saramaccan (dont toutefois je ne connaissais pas encore les règles de perturbation) je ne pouvais encore lire sans faute, à la cadence d'une conversation normale, des textes que j'avais écrits moi-même. J'ai d'ailleurs constaté ensuite combien les progrès sont rapides dès que l'on connait ces règles.

Le principal problème posé au linguiste dans cette étude est, je le crois, celui de l'analyse syntaxique d'une langue tonale comme celle que j'ai décrite. A mon avis, l'analyse syntaxique des langues créoles est particulièrement difficile parce que la valence syntaxique du mot dans la proposition ne se révèle pas par la forme du mot mais exclusivement par sa position dans la proposition et les possibilitiés de ses combinaisons avec d'autres mots. C'est pour cela que, pour moi, il était tout particulièrement important de pouvoir établir une relation bien nette entre le ton et la grammaire. Il me semble que, pour une grande partie, nous pouvons baser notre analyse syntaxique du Saramaccan sur le comportement tonal des mots dans la proposition.

Amsterdam

WILLIAM R. SCHMALSTIEG

Baltic Nominal Systems

In a recent article Edward Stankiewicz has examined the relationships between the occurrence of prosodic features and consonantal alternations in the Slavic noun paradigms.[1] Stankiewicz (p. 186) finds that, ". . . in Slavic nominal systems with productive accentual alternations, consonantal alternations are absent or unproductive; consonantal alternations are, on the other hand, productive in systems without accentual alternations."

In the Russian noun declension, where there is a rather complex system of stress alternation, there remain only a few words which have any kind of consonantal alternation, e.g. Russ. /čort/ 'devil', /mudˈo/ 'testicle', /sasˈet/ 'neighbor', /kal͡ʲena/ have the nom. pl. /čʲert͡i, mˈud͡i (or /mud͡ʲe/), sasˈed͡i, kal͡ʲen͡i/ with an alternation between stem final palatalized and unpalatalized consonant. There are also some residual forms such as /ˈuxo/ 'ear', /ˈoko/ 'eye' (archaic) with the plurals /ˈoči, ˈuši/ respectively. The accentual oppositions of Russian are mostly redundant, but in some cases they do sharpen the polarity between the singular and the plural, e.g. nom. sg. /dušˈa/ 'soul', /slʲova/ 'word' as opposed to the nom. pl. /dʲuši slavˈa/ respectively.[2]

In Polish the stress is fixed on the penultimate syllable. The nominal declension, however, is rich in consonantal alternations. For example, in the nominative plural of personal nouns we find alternations between (1) velar and non-velar, (2) unpalatalized and palatalized consonants, (3) dental and palatal consonants, e.g. (1) /fizik/ 'physicist', /kolega/ 'colleague', /mńix/ 'monk' have the nom. pl. /fizici, koleʒi, mńiśi/, (2) /b͡iskup/ 'bishop', /snop/ 'snob', /agronom/ 'agronomist' have the nom. pl. /b͡iskup͡i, snob͡i, agronom͡i/, (3) /idjota/ 'idiot', /sǫśat/ 'neighbor' have the nom. pl. /idjoći, sǫśeʑi/.[3]

Stankiewicz examines the other Slavic languages also, but there is no need to repeat any more here. It is clear that in the nominal systems of the

[1] "The Consonantal Alternations in the Slavic Declensions," *Word*, XVI (1960), 183–203.

[2] *Ibid.*, p. 189.

[3] *Ibid.*, p. 198. Cf. also Charles Bidwell, "The Morphophonemics of Polish," *General Linguistics*, II (1957), pp. 71–99.

Slavic languages the accentual and the consonantal alternations are in an inverse relationship and that ". . . the productivity of one type is incompatible with the productivity of the other."[4]

It is interesting to note that such an inverse relationship of the productivity of the consonantal and accentual alternations is to be found in the nominal systems of the two living Baltic languages, Lettish and Lithuanian.

In the declension of the Lithuanian noun only two consonants, /d͡/ and /t͡/, are subject to morphophonemic alternation. These consonants alternate with /dž͡/ and /č͡/ respectively.

Stem final /dž͡/ or /č͡/ is found in every case of those etymological **jo*-stem nouns which have the nominative singular in /-is/ or /-i:s/ except the following: the nominative singular, e.g. /žo˜:d͡is/ 'word', /šal˜t͡is/ 'cold', the accusative singular, e.g. /žo˜:d͡i:, šal˜t͡i:/, the locative singular /žo˜:d͡i:je, šal˜t͡i:je/, the vocative singular /žo˜:d͡i, šal˜t͡i/. In the etymological **jo*-stems which have the nominative singular in /-es/, e.g. /sv͡e˜:č͡es/ 'guest', the accusative singular is /sv͡e˜:č͡e:/, but all of the other morphophonemic alternations are the same as those of /žo˜:d͡is, šal˜t͡is/, etc. We find /dž͡, č͡/ as stem final alternants of /d͡, t͡/ in the dative singular of the etymological *i*-stems (e.g. /šir˜d͡is/ 'heart' has the dat. sg. /šir˜dž͡ei/) and in the genitive plural of the *ė*-stems and some of the etymological *i*-stems (e.g. /b͡i˜t͡e:/ 'bee' and /šir˜d͡is/ 'heart' have the gen. pl. /b͡i˜č͡u:, šir˜dž͡u˜:/ respectively). An alternate instrumental singular from the **jā*-stems is sometimes found in the *i*-stems, e.g. Lith. /šir˜dž͡e/ 'heart' in place of the etymological /šir˜d͡imì/.

Although the dual is not much used in modern standard Lithuanian, it is necessary to mention that words such as /valdž͡ė̀/ 'government' have the nom. acc. dual /(dv͡i) valdì/ '(two) governments'. It is interesting to note that Kurschat lists the dual *waldźì* where the *-dź-* (=/dž͡/) may have been resubstituted by analogy, thereby removing the original morphophonemic alternation.[5]

In Lithuanian one of the nominal accent classes (generally termed 'class I'), just as in Russian, has a constant stress on the stem syllable throughout the declension, e.g. /v͡í:ras/ 'man', /bró:l͡is/ 'brother', /l͡iepa/ 'linden tree', /ánt͡is/ 'duck', etc.[6] On the other hand, there are three other

[4] Stankiewicz, *op. cit.*, p. 202.

[5] *Grammatik der littauischen Sprache* (Halle, 1876), p. 172.

[6] Although I use the conventional signs -'- (acute) and -~- (circumflex) to mark the Lithuanian stress, phonemically the basis is not musical tone, but comparative amplitude. The acute signifies greater amplitude on the first part of the syllable (*-á-*=/áa/ or /á:/), whereas the circumflex signifies greater amplitude on the second part of the syllable (*-ã-*=/aá/ or /a!/). References: A. Laigonaitė, "Dėl lietuvių kalbos kirčio ir priegaidės supratimo," *Kalbotyra* I (1958)=Vilniaus Valstybinis V. Kapsuko Vardo

accent classes in which the stress appears on the root in some cases and on the declensional ending in others. Rules for these alternations can be found in any grammar of Lithuanian and they do not need to be repeated here.[7] One can point out, however, that for accent classes II, III and IV, certain cases are distinguished only by the position of the stress. In the etymological *ŏ*-stems we find the following locatives singular: (accent class II) /pir̃štè/ 'finger', (III) /kal̃nè/ 'hill', (IV) /dìeṽè/ 'God' as opposed to the vocatives /pir̃šte, kál̃ne, dìeṽe/ respectively.[8] In the etymological **jŏ*-stem (or **(i)jŏ*-stem in traditional treatments) nouns we find the accusative singular (III) /árkl̃i/: 'horse', (IV) /pir̃šl̃i:/ 'match-maker' as opposed to the vocatives singular /arkl̃i:, piršl̃i:/. There are also alternate locatives singular /arkl̃i:, piršl̃i:/ beside the regular /arkl̃i:jè, piršl̃i:jè/.

In the nominative plural of the **ā*-, **jā*- and *ė*-stems the stress is never on the ending, cf. nom. pl. (I) /l̃iepo:s/ 'linden trees', /kó:jo:s/ 'feet', /g̃érv̂ė:s/ 'cranes'; (II) /rañko:s 'hands', /dva:s̃o:s/ 'souls', /žẽ:m̂ė:s/ 'lands'; (III) /gálvo:s/ 'heads', /g̃iesm̂ė:s/ 'hymns'; (IV) /žiẽmo:s/ 'winters', /vaĺdžo:s/ 'governments', /pe:l̃ės/ 'mice'. In accent classes I and II the nominative plural and the genitive singular are homonymous, but in accent classes III and IV the stress is always on the genitive singular ending, e.g. /galvo:s, g̃iesm̂ė:s, žiemo:s, valdžo:s, pel̃ė:s/ respectively. Here the genitive singular and nominative plural are distinguished only by the position of the stress. For the **ā*- and **jā*-stems in accent class III the instrumental singular is homonymous with the vocative singular e.g. /gálva/ as distinguished from the nom. sg. /galvà/, whereas in classes II and IV the instrumental singular is homonymous with the nominative singular, e.g. (II) /rankà, dvasè/, (IV) /žiemà, valdžè/ as opposed to the voc. sg. /rañka, dva:s̃e, žiẽma, vaĺdže/. In accent classes II and IV of the *ė*-stems the instrumental singular has the stress on the ending, e.g. (II) /šv̂entè/ 'holiday',

---

Universitetas Mokslo Darbai XXVI, pp. 71–100; Jerzy Kuryłowicz, *L'accentuation des langues indo-européennes* (Cracow, 1952), p. 199.

[7] Cf. also *Dabartinės lietuvių kalbos žodynas*, ed. J. Balčikonis, et. al. (Vilnius, 1954), pp. xi–xiv.

[8] The accentuation /dìev̂è/ in the vocative is also known. Frequently the voc. sing. /tè:̃ve/ is cited as a form which shows a metatony of /tė:vas/ 'father' (e.g. by K. Būga, *Lietuvių kalbos žodynas* (Kaunas, 1924), p. XXXI). I am indebted to Prof. Anthony Salys of the University of Pennsylvania for pointing out to me that this is probably a dialect form with a secondary circumflex which is the result of the withdrawal of the accent from a form with the stress on the final syllable. Names with three or more syllables may have various types of accentuation, e.g. the vocative of *Antanas* 'Anthony' might be either /anta:̃n, añtan/ as well as /antãne/, cf. J. Endzelin, *Baltų kalbų garsai ir formos* (Vilnius, 1957), p. 106. This is the Lithuanian translation of his *Baltu valodu skaņas un formas* (Riga, 1948).

(IV) /žẽmè/ 'earth', whereas the vocative has the stress on the stem, e.g. /švẽnte, žẽ:me/ respectively.

Although the dual is not important in the modern language, it could be noted that for all nominal stems in accent classes III and IV the dative dual differs from the instrumental dual only by the position of the stress, cf. the datives dual (III) /giesmé:m/ '(two) hymns', /širdìm/ '(two) hearts', (IV) /svečém/ '(two) guests', /žiemó:m/' (two) winters' as opposed to the respective inst. dual /giesmė̃:m, širdim̃, svečem̃, žiemo:̃m/.

Certain cases, such as the accusative and dative singular are marked by the additional feature of always having the stem stress in all accent classes and for all etymological stem classes.

Sometimes in Lithuanian dialects the nominative plural is distinguished from the vocative plural by the position of the stress, cf. the nom. pl. /vaikaĩ/ 'children', /bernaĩ/ 'lads' beside the voc. pl. /vaĩkai, beĩnai/.[9]

A difference in the position of the stress may denote a semantic difference, e.g. the noun /tė́:vas/ 'father' has the nom. pl. /tė́:vai/; 'fathers' but the concept 'parents' is rendered by the same form except that the stress is on the final syllable, i.e. /tė:vaĩ/.

As opposed to Lithuanian, where in the nominal declension only two consonants are subject to morphophonemic alternations, in Lettish eleven consonants and one consonant cluster undergo morphophonemic alternation in the etymological **jŏ*-stems and consonant stems. (Historically these morphophonemic alternants are the result of a combination of consonant plus /j/ before a non-front vowel.) Since the only essential distinction between the consonant and **jo*-stems is that the former have *-s* in the genitive singular, whereas the latter have *-a*, they are grouped together in the Academy Grammar as the second declension.[10] In standard Lettish the mutated consonants are found in the genitive singular (if it has the ending *-a*) and all of the cases of the plural. Essentially, however, the consonantal alternation appears to be a morphological device for sharpening the polarity between the singular and the plural. In the following examples of consonantal alternation the nominative singular is listed first, then the genitive singular, then the nominative plural: (1) *p–pj* /u:pis, u:pja, u:pji/ 'owl'; (2) *b–bj* /gùlbis, gùlbja, gùlbji/ 'swan'; (3) *m–mj* /kurmis, kurmju, kurmji/ 'mole'; (4) *v–vj* /cìrvis, cìrvju, cìrvji/ 'axe'; (5) *t–š* /zutis, zuša, zuši/ 'eel'; (6) *d–ž*/briẽdis, briẽža, briẽži/ 'deer'; (7) *s–š* /lasis, laša, laši/ 'salmon'; (8) *z–ž* /nazis, naža, naži/ 'knife'; (9) *n–ņ* /li:nis, li:ņa li:ņi/ 'tench'; (10) *l–ļ* /bra:lis, bra:ļa, bra:ļi/ 'brother'; (11) *r–ŗ* /stù:ris, stù:ŗa,

[9] Cf. Endzelin, *op. cit.*, p. 107.

[10] *Mūsdienu latviešu literārās valodas gramatika*, ed. A. Bergmane et al. (Riga, 1959), p. 412.

stù:ŗi/ 'corner'; (12) *st–šķ* /strupastis, strupašķa, strupašķi/ 'bob-tailed dog'.[11]

Almost all of these morphophonemic alternations are also found in the genitive plural of the *e*-stem nouns, e.g. nom. sing. /sliẽde/ 'rail', /là:se/ 'drop', /lùoksne/ 'sheet' which have the gen. plur. /sliẽžu, là:šu, lùokšņu/ etc.[12] Certain of the etymological *i*-stems have the mutated consonant in the genitive plural, e.g. /uguns/ 'fire', /kliņts/ 'cliff' have the gen. plur. /uguņu, kliņšu/ respectively.

The initial intensity stress of Lettish can be considered a phenomenon of juncture and need not concern us here.[13] In the stressed (i.e. initial in most cases) syllable the Lettish syllable intonations are: (1) the falling tone (rising and then falling), (2) the Dehnton (rising and then level), (3) the Stosston (rising, then a glottal catch or diminution of amplitude and then a fall).[14] Outside of the stressed syllable all of the intonations are falling, but the Dehnton levels out after the fall and the Stosston has a broken fall.[15] According to the Academy Grammar (p. 67) the three intonations are sharply distinguished only in the stressed syllables and in unstressed syllables the intonations are very unstable. In contemporary Lettish the three basic intonations are distinguished only in some of the central dialects. Elsewhere there are two basic intonations varying in quality depending upon the region.

In the nominal inflection the syllable intonation has only a very limited morphological function.[16] A root syllable Dehnton is replaced by the falling tone in the vocative, e.g. the vocatives of /bra:̃lis/ 'brother', /ma:̃te/ 'mother', /siẽva/ 'woman' are respectively /brà:li, mà:t, sìev/. No other cases in Lettish are distinguished by the type of intonation of the root syllable.

[11] The examples are taken from the Academy Grammar (cited in fn. 10), p. 413. According to this grammar (p. 414) the mutated consonant may be found in the dative singular also, e.g. /u:̃pjam, gùlbjam . . ./, but it is not widely used.

[12] In some words (cf. the Academy Grammar cited in fn. 10, p. 419) this morphophonemic rule is not observed: (1) in case of a possible semantic confusion, e.g. the genitive plural of /mute/ 'mouth' is /mutu/, not /mušu/ which is the genitive plural of /muša/ 'fly'; (2) in compound and derivative words with the final syllable in *-de*, *-te*, e.g. /piecgade/ 'five-year plan', gen. plur. /piecgadu/; (3) many foreign words, e.g. /epizode/ 'episode', gen. plur. /epizodu/, although here vacillation is found with some words, e.g. /bala:de/ 'ballad', gen. plur. either /bala:žu/ or /bala:du/.

[13] Eric Hamp, "The Phonemic Structure of Lettish Syllables," *Die Welt der Slaven*, Vol. III (1958), p. 248.

[14] R. Ekblom, *Die lettischen Akzentarten* (Uppsala, 1933), p. 21.

[15] Hamp, *op. cit.*, p. 251.

[16] Differences of intonation are important in derivation, however, cf. the Academy Grammar, p. 69.

Endzelin's grammar lists the Dehnton as the intonation of the locative singular and dative and instrumental plural endings, e.g. loc. sing. /bra˜li˜/ 'brother', /cìlve˜ka˜/ 'man', /siẽva˜/ 'woman', /ma˜te˜/ 'mother', /avi˜/ 'sheep', /akmeni˜/ 'stone' and the dat. inst. plur. /bra˜liẽm, cìlve˜kiẽm, siẽva˜m, ma˜te˜m, avi˜m, akmeņiẽm/ respectively.[17] Endzelin also lists the Stosston as the intonation of the locative e.g. /bra˜luôs, cìlve˜kuôs, siẽva˜s, ma˜te˜s, avi˜s/. On the other hand the Academy Grammar lists no intonations at all in the sections devoted to morphology. The reason given (p. 68) is that up to the present there have been no investigations concerning the types of intonation in the literary language.

Thus in the Lettish nominal declension the intensity stress has no functional load and the syllable intonation has a very limited functional load, whereas the consonantal alternations are quite complex.[18] This is the reverse of the situation in Lithuanian where there are four different classes of nominal accentuation and only the voiced and unvoiced dental stops are subject to morphophonemic alternations.

One might ask the historical reason for the larger number of consonantal alternations in Lettish. In Common Baltic there was a contrast of palatalized versus unpalatalized consonants before non-front vowels, but no contrast before front vowels.[19] Phonemically the same thing is true in modern Lithuanian, although the Lithuanian situation does not quite reflect that of Common Baltic. In Lithuanian the groups */nja, mja, lja . . ./ have fallen together with */ne, me, le . . ./ thereby erasing one phonemic distinction, although (in modern standard Lithuanian) */dja, tja/ have become /d͡ẑe, t͡ŝe/ instead of merging with */de, te/. In modern standard Lettish, on the other hand, the phonemic distinction between all etymologically palatalized consonants is now found before front vowels as well as back vowels.

In Lettish, then, the creation of morphophonemic complexity in the consonantal alternations was accompanied by a corresponding decrease in the prosodic morphophonemic alternations, either through analogy

[17] *Latviešu valodas gramatika* (Riga, 1951), passim.

[18] The etymologically stem-stressed nouns should have a locative singular and plural with the Dehnton (phonologically Lett. /pupa˜/ = Lith. /pùpoje/ 'in the bean', accent class II, etymologically stem-stressed), whereas the etymologically end-stressed nouns should have a locative singular and plural with the Stosston (phonologically Lett. /vagâ:/ = Lith. /vago:jè/ 'in the furrow', accent class 4, etymologically end-stressed). In each case an analogical generalization has taken place such that in standard Lettish in the singular we find the Dehnton (Lett. /pupa˜, vaga˜/) and in the plural the Stosston (Lett. /pupâ:s, vagâ:s/), cf. Endzelin, *Latv. val. gram.*, p. 42.

[19] This situation arose when the Indo-European groups */Cje/ and */Ce/ merged as */Ce/ (or */t͡Ce/?) in Common Baltic.

(cf. fn. 18) or through phonological development (i.e. the creation of a phonetic intensity stress on the initial syllable). Thus, similarly to the Slavic languages, there is an inverse relationship between the number of consonantal and the number of accentual alternations in the nominal systems of the living Baltic languages.

Lafayette College
Easton, Penna.

R. B. LEES

Grammatical Analysis of the English Comparative Construction

No standard handbook of English grammar fails to describe for the reader the difference between those adjectives which are "compared" by means of the *-er* and *-est* suffixes on the one hand, and those polysyllabic adjectives, or adjectivals, which require the use of a preceding *more* and *most*. Examples will always be given of sentences containing the *as-much-as* and the *more-than* constructions. And compendious works, such as those of Jespersen and Poutsma, will also provide numerous examples of other more or less related constructions in *as* and *than*.[1]

But there is also a number of previously unmentioned restrictions and peculiarities characterizing these, or some of these, constructions which demand explanation. In some cases a handbook will offer a historical "explanation" of a particular structure, should one be available, but since all of these constructions may be productively derived *ad libitum* it is reasonable to demand of the grammarian that he explain *how* an unschooled speaker of English manages to do this.[2] To my knowledge, the only satisfying way to accomplish this is to formulate a more or less complex, ordered set of grammatical rules which correctly enumerates all and only the well-formed sentences containing these constructions and which fulfills certain strict but reasonable requirements of grammatical analysis. For example, the rules offered must fit in well with an adequate total grammar of English sentences; the way in which each expression is derived by means

[1] See, for example, Otto Jespersen, *A Modern English Grammar* (Part VII: Syntax), London: George Allen & Unwin Ltd., 1954 and 1958, Chapters X and XI: Comparison; G. O. Curme, *Syntax*, Boston: D. C. Heath and Co., 1931, pp. 498–508, and for various similar constructions pages 294–308.

[2] That is, one can reasonably demand two kinds of linguistic explanations: (a) how a speaker is able correctly to construct an unlimited number of new comparative expressions, and (b) how the language in which he constructs them evolved into its present state. Clearly, the historical explanation (b) presupposes the descriptive explanation (a), for to specify how the rules of a language have changed through time one must first be able to specify the rules.

of the rules must automatically yield an exact specification of its grammatical structure; and these rules must be completely explicit, mechanically applicable, and free of dependence upon intuition or special art.[3]

The most obvious constraints on the free generation of comparative constructions are, of course, precisely those least likely to be pointed out, for they appear naturally to fall under common-sense semantic considerations. Yet, upon closer examination, we can see that many regularities in these conditions are easily explained on purely grammatical grounds. (Let me remark parenthetically that it is quite reasonable to welcome success in explaining any apparently semantic regularity on the basis of grammar as a concrete result of linguistic science, for it is precisely the semantic features of sentences which are in such bad need of elucidation and the grammatical features which can so easily be studied independently.)[4]

We shall begin with a rather commonplace observation, but one the explanation of which already requires us to formulate a set of rules for deriving comparative constructions from simpler sentences. Notice the following facts:

(1) The boy is amazed.
(2) The man is amazed.

but not:

(3) *The table is amazed.

and, correspondingly, also not:

(4) *The boy is more amazed than the table.

although:

(5) The boy is more amazed than the man.

Or, similarly, we have the following triplet:

(6) The meeting is adjourned soon.
(7) *John is adjourned soon.
(8) *The meeting is adjourned sooner than John.

Note, incidentally, that there is nothing inherently wrong with the particular sequence of words chosen in (8), for we have:

(9) The meeting is adjourned sooner than John expected.

[3] The descriptive explanation consists then simply in imputing the formulated rule to the speaker, or to his knowledge of the language.

[4] See R. B. Lees, "The Grammatical Basis of Some Semantic Notions," *Eleventh Annual Round Table Conference*, Georgetown University, 1960 (proceedings to appear).

It is just that we know intuitively that if the sequence in question [i.e., **(8)**] is to be a sentence, then it must have a grammatical structure very similar to that of another type, exemplified by the equally impossible:

(10) *The meeting is adjourned sooner than John is adjourned.

and it is the latter part of *this* sequence which has caused the difficulty.

If we could account for the fact that the pseudo-sentence:

(11) *John is adjourned.

is ungrammatical, then we could account in the same way for the impossibility of **(10)** and therefore also perhaps of **(8)**. Of course, to do so would further require us to connect the two in some formal way. But this is exactly what a transformational generative grammar of English enables us to do.[5]

Now it has already been pointed out in many other places that if we are to formulate a reasonably simple grammar for English, then we shall wish to employ exactly the same apparatus to effect these two connected restrictions simultaneously, and that this is alone sufficient to motivate our formulating the grammar in such a way that the longer sentence-type **(10)** be derived *from* the shorter **(11)**, perhaps together with some other underlying source-expression.[6] Before attempting such a formulation, however, let us observe some other peculiarities.

Notice next that there are certain sentences which can be understood in two different ways, i.e., which are structurally ambiguous:

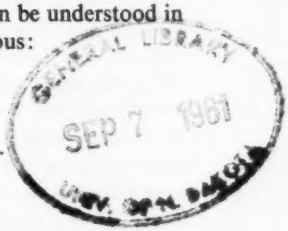
(12) I speak Spanish as well as French.

That is:

(13) I speak Spanish as fluently as I speak French.

or:

(14) I speak Spanish in addition to French.



Taking a truly hard-headed position, we may have to recognize still a third interpretation:

(15) I speak Spanish as well as [Mr.] French does.

Whatever derivation we decide to use to explain comparative sentences, we

[5] For the latest summary of the arguments for a transformational extension of the usual constituent-structure expansion-rule model of grammars, see N. A. Chomsky, "On the Notion 'Rule of Grammar'," *Proceedings of the Symposium on the Structure of Language and its Mathematical Aspects*, American Mathematical Society (Providence, R.I.), April, 1961.

[6] *Ibid.*, also see R. B. Lees, *The Grammar of English Nominalizations*, Bloomington: Indiana University Research Center in Anthropology, Folklore, and Linguistics, Publication Number 12, July, 1960, (*International Journal of American Linguistics* XXVI, Number 3, Part II), especially pp. xviii–xix.

shall want it also to cast light on why the sentence in question (12) is ambiguous, in fact even triply ambiguous.

Again, it has already been noted that a grammatical description in which some sentence-types are derived from others can easily serve to explicate ambiguities; we require that on independent grounds it be expedient to formulate the rules so as to provide at least two different ways in which one particular sequence of morphemes can be derived.[7] In this particular case, we aim at rules for generating comparatives which should yield the given sentence (12); but we also require that some other rules yield that very sequence. In particular, we shall wish the sentence when construed like (13) or like (15) to be some kind of comparative, but when construed like (14) to be a sort of conjunction.

Ambiguities are cases in which one sentence has two different grammatical structures. But there are also cases in which two different sentences which appear to have the same structure on the surface must be assigned contrasting underlying syntactic constituent-structure.[8] For example:

(16) He comes as frequently as John.
(17) He comes as frequently as every day.

are intuitively quite different, although their morpheme sequences are similar. And there are also forbidden examples, such as:

(18) *He comes as frequently as six feet.
(19) *He jumps as far as every day.

Here we might say that there are certain classes of adverbials, that the constituent inside of the comparative construction is an adverbial, that the final constituent is also an adverbial, and finally that these two adverbials are constrained to choice from the same class.

A minor question impinging upon matters of good style is often mentioned in handbooks, namely the choice of subjective or objective form for pronouns in final position after comparative expressions, as in:

(20) I know him better than she.
(21) I know him better than her.

It is also pointed out that these latter two cases can be in contrast, parallel to:

(22) I know him better than she does.
(23) I know him better than I do her.[9]

[7] See N. A. Chomsky, "Three Models for the Description of Language," Institute of Radio Engineers, *Transactions on Information Theory*, IT-2 (1956), p. 118.

[8] As in the cases discussed in N. A. Chomsky, *Syntactic Structures*, 's-Gravenhage: Mouton & Co., 1957, pp. 73–5, 81–3, 89–90; or in R. B. Lees, "A Multiply Ambiguous Adjectival Construction in English," *Language*, XXXVI, 1960, 207–221.

[9] As in Curme, *op. cit.*, page 303.

Thus, if the first two sentences **(20)** and **(21)** are derived *from* the latter two we shall be able to explain the contrast, but we shall then have to formulate exactly how the reductions are to take place. For example, we cannot say simply that the final verb may be first replaced by the pro-verb *do* and then may be deleted entirely, else we should permit:

(24) I know him better than they say she does.→
(25) *I know him better than they say she.

Another constraint which must be built into the rules we seek to formulate is on the appearance of the negative particle. Notice that the first constituent of the comparative construction is freely affirmative or negative, but that the second constituent cannot be negative:

(26) I know him better than she does.
(27) I don't know him better than she does.
(28) *I know him better than she doesn't.

Notice furthermore that this constraint is true also for the so-called "negative verbs," for we also have no:

(29) *I know him better than they denied she does.[10]

While we are considering this constraint on the appearance of negative in the comparative construction it is also appropriate to note another peculiarity. Like negative sentences, questions, subjunctive expressions, and others, the comparative construction may be called "affective," following E. S. Klima.[11] These so-called affectives may contain *ever* and also *any* for *some*, as well as other special expressions:

(30) I haven't ever seen any.
(31) I have seen some.
(32) *I have ever seen any.

Similarly, with comparatives:

(33) It is as good as I have ever seen.
(34) It is as good as any I have seen.
(35) It's as good as ever.

Furthermore, there is a special negative pre-modifier for the comparative:

(36) I'm not taller than he.
(37) I'm no taller than he.
(38) *He is no tall.

[10] The "Negativity" of certain verbs was first pointed out to me by Professor E. S. Klima of the Massachusetts Institute of Technology in a paper read before the Seminar on Transformational Grammar at University of Pennsylvania, 1959.

[11] A study of his so-called "affective" expressions in now in preparation.

Before formulating the actual rules for enumerating such expressions it is essential that we decide, at least tentatively, what kind of constituent-structure the comparative constructions are going to be said to have. We have already presented evidence in favor of the view that

(39) He is as tall as she.

is to be derived from:

(40) He is as tall as she is (tall).

But the same derivation is not available for the minimal

(41) He is as tall as that.

It would seem reasonable to regard the material which follows the second *as* in **(39)** through **(41)** as a kind of adverbial, parallel to the *that* of:

(42) He's that tall.

from which we might also later derive the as yet unanalyzed:

(43) How tall is he?
(44) That's how tall he is.

(though perhaps not the anomalous:

(45) That's all the taller he is.)

Notice parenthetically, however, that it is not exactly parallel to the pre-adjectival modifier in:

(46) He's *six feet* tall.

since such a "quantifier" may occur only before one of a special group of quantifiable adjectives such as *tall*, *wide*, *deep*, *old*, etc., but there is no such pre-adjectival modifier possible before other adjectives, as in:

(47) He's that warm.
(48) It's that new.

Now, the *that* of **(41)** would seem to be a *pro*-adverbial standing for any of the indefinitely long expressions which can follow *as* in such a sentence. The *that* of **(42)**, on the other hand, appears to be a kind of pre-adverbial or pre-adjectival adverbial modifier which, though differing from *very*, *quite*, *rather*, etc., will be very useful now in deriving questions and comparatives in the following way: by operating upon the *that* as adverbial with the so-called WH-transformation we obtain:

(49) How tall he is . . .

Or, had the sentence already been questioned, we would have derived:

(50) Is he that tall?→
(43) How tall is he?

And secondly, considering the *as . . . as* and the *more . . . than* expressions to be (comparative) operators like the WH, and permitting them to replace the *that*, even though discontinuously, we obtain:

(42) He is that tall. →He is as tall as . . .
or:
→He is more tall than . . .

Of course, later morphophonemic rules which ensure the proper phonemic shape for comparative adjectives will also include rules to convert such expressions with mono-syllabic adjectives as *more tall* into the appropriate inflected form *taller*.

Thus, we have indicated an analysis of comparatives, except for special conditions and constraints which must be imposed, though it is not yet completely clear what the form of the other underlying sentence must be. To begin with, we might suppose that the latter has the simplest possible form, as in:

(42) He is that tall.
(51) } →He is as tall as I am.
(52) I am tall.

and we suppose further that subsequent rules permit certain reductions of the transform to yield:

(53) He is as tall as I.

and even certain stylistic developments to yield the colloquial:

(54) He is as tall as me.

But if we are to have the most uniform rules, and if we are at the same time to account for such examples as:

(55) He is as tall as I am wide.

then we must suppose that the transform contains the entire constituent-sentence (as I call the second source-sentence in a generalized grammatical transformation rule).[12] This produces awkward expressions which must be reduced, but the conditions for these reductions may be easy to formulate. Note, incidentally, that we do not shy away from introducing further complications of this kind since they affect only the obligatory rules of the

[12] See Lees, *Nominalizations*, p. 55.

grammar, permitting greater simplifications in the optional component.[13] The desired reduction need only delete the adjective whenever it is identical to the preceding adjective, and the derivation would now be:

(42) He is that tall.
(56) → He is as tall as I am tall.
(51) I am tall.

(52) $\xrightarrow{ob}$ He is as tall as I am.

(53) $\xrightarrow{op}$ He is as tall as I.

(41) $\xrightarrow{op}$ He is as tall as that.

Let me remark also why we choose to include the *that*-constituent in the matrix-sentence (or first source-sentence of the transformation rule).[14] It seems that the constituent-structure of transforms may most easily be derived from that of the source-sentences by a uniform process of substitution: linguistic theory need say only that the constituent-structure of a component of a transform is simply that of the underlying constituent which it has replaced.[15] In our case, the constituent-structure of the component *as tall as I am* is Adjective, since it replaced the adjectival constituent *that tall* in the matrix-sentence; second, the constituent-structure of *as . . . as I am*, which happens to be phonemically discontinuous, is Attributive Adverbial, since it replaced the adverbial *that* in the matrix-sentence. We have now only to say how the latter adverbial had obtained this constituent-structure.

But this is easy . . . we need only permit the morpheme *that* to be generated along with *very*, *quite*, *rather*, etc. in the constituent-structure component of the grammar by means of ordinary expansion-rules together with its following adjectival. For example, in the Copula-type sentence of the form:

$$\text{Nom} + \text{Aux} + \text{be} + \text{Pred}$$

we permit the predicate Pred component to be expanded by a rule as

$$\text{Pred} \rightarrow \left\{ \begin{matrix} \text{Nom} \\ \text{Loc} \\ (\text{Adv}_a) + \text{Adj} \end{matrix} \right\}$$

where Nom is a nominal, Loc is a "locative" place-adverbial, and $\text{Adv}_a + \text{Adj}$ is an adjective modified by a preceding attributive adverbial,

[13] See Lees, *Nominalizations*, p. 32.
[14] See footnote 12.
[15] Lees, *ibid.*, p. 50, footnote 24; also Chomsky, "Rule of Grammar," section 3.

one of which is the morpheme *that*.[16] This constituent-structure Adv_a is then transmitted through the later transformation to the substituent *as . . . as I am tall*, or *more . . . than I am tall*, which replaces the *that*.

From the meaning of the transform-sentence we might wonder whether the adjectival from the second, or constituent-, sentence does not also contain a similar attributive adverbial. One reason this might be desirable is that it would help us to specify more rigorously what can be the form of the second source-sentence, for clearly we cannot have any arbitrary sentence whatever in that position, as in:

(42) He is that tall.
(57) } *He is as tall as I was a new arrival.
(58) I was a new arrival.

although it is remotely possible that we shall wish to include as grammatical:

(42) He is that tall.
(59) } → He is as tall as I was new an arrival.
(60) I was that new an arrival.

We shall gain more information on this question by looking now at other kinds of examples of the comparative construction.

We note that comparatives may also be constructed with adverbial expressions, as in:

(61) He walks as slowly as I do.

Again, in these adverbial versions we are led to formulate the transformation so as to yield the full second source-sentence with subsequent obligatory and optional reductions, for we might also have such examples as:

(62) He sings as badly as she plays well.

which we can presume to be derived from:

(63) He sings that badly.
(64) She plays well.

Note also, incidentally, that there is no necessity to reduce the second verb when the two underlying subjects are different:

(65) He sings as badly as she sings. $\xrightarrow{op}$
(66) He sings as badly as she does. $\xrightarrow{op}$
(67) He sings as badly as she.

[16] This is then a slight extension, or improvement, of the Constituent-structure expansion-rules given in Lees, *ibid.*, p. 13.

In addition to these examples produced by a process which turns on adverbial expressions in exactly the same way as did the adjectival examples presented before, there are also others of the form:

(68) The attacks come as frequently as once a day.

which of course cannot be construed in the same way as:

(69) The attacks come as frequently as the retreats.

Here the discontinuous *as . . . as once a day* component appears to be parallel in function to the former *as . . . as I am*; that is, it appears to be an adverbial modifier of *frequently*, as though derived from an expresion *that frequently*, parallel to *that tall*. Accordingly, we may derive:

(70) The attacks come that frequently.
(71) The attacks come once a day. → (T)

(72) The attacks come as frequently as the attacks come once a day. →(obligatorily)

(68) The attacks come as frequently as once a day.

And again, the reason we choose to include in the transform for further reduction the obligatorily deleted subject and verb is that there are other examples in which, when the subject and verb do *not* repeat those of the preceding source-sentence component, the latter are *not* elided:

(73) The attacks come as frequently as the retreats came rarely.

From these examples it would seem that the transformational rule in question can be said to apply just in case the two source-sentences contain each an appropriate *Adverbial*. But in the first set of derivations we formulated for the adjectival case there was no adverbial constituent in the second source-sentence (**52**). Perhaps it is then correct to require that there be one, in the following way:

(42) He is that tall.
(74)
(75) I am that tall. → He is as tall as I am that tall.

(56) $\xrightarrow{ob}$ He is as tall as I am tall.

(52) $\xrightarrow{ob}$ He is as tall as I am.

Before summarizing in formal rules these constraints on the comparative construction, let us first examine in greater detail the particularities of the subsequent reduction-transformation rules which will be required. First of

all, the attributive adverbial *that* which appears in the latter part of the transform must be deleted in all cases, for we have no:

(76) *He is taller than I am that wide.

Second, when the adjectival constituent in the second part exactly repeats that of the underlying matrix-sentence, then it too is deleted:

(77) He is taller than I am wide.
(78) He is taller than I am.
(79) He is taller than I.
(80) *He is taller than I am tall

Third, whenever the second verb exactly repeats the first, then it too may be deleted, but only optionally:

(81) He eats faster than I eat.
(82) He eats faster than I.

(83) He can eat faster than I can eat.
(84) He can eat faster than I can.

(85) He is eating faster than I am eating.
(86) He is eating faster than I am.

From these examples we also see that the preceding auxiliary can remain *in toto*; we are therefore led to regard the first case (**82**) as a secondary reduction from:

(87) He eats faster than I do.

for the pro-verb *do* can then be taken as the normal carrier of the Tense morpheme for the case where there is no other member of the Auxiliary component present, as in Negative and in Questions.[17]

We must then go on to specify the conditions under which various members of the AUX can be deleted and also when the entire Verb Phrase is reduced to zero. First, we note that any number of AUX members may be retained by deletion in order from the right-hand side:

(88) He would have been talking louder than I would have been talking.
(89) He would have been talking louder than I would have been.
(90) He would have been talking louder than I would have.
(91) He would have been talking louder than I would.
(92) He would have been talking louder than I.

[17] Lees, *ibid.*, p. 35–6; Rule (T41*), p. 49; and references to Chomsky therein.

Second, we must note that the last reduction to zero-verb-phrase is not always permitted, for we have:

(93) He says that she talks more loudly than they say she does.
(94) He says that she talks more loudly than they do.
(95) He says that she talks more loudly than they.
(96) *He says that she talks more loudly than they say she.

That is to say, the final verb-phrase, reduced to a phrase in *do*, *have*, *be*, or a modal such as *can*, is not reducible to zero if it is within a subordinate clause of the constituent-sentence; only the verb-phrase of the main clause of the underlying constituent-sentence may be completely deleted.

We are now finally ready to formulate a set of rules to generate such constructions.[18]

$$GT.\quad \left.\begin{matrix} X+that+A+Y \\ Z+B+W \end{matrix}\right\} \rightarrow X+\left\{\begin{matrix} as+A+Y+as \\ more+A+Y+than \end{matrix}\right\}Z+B+W$$

where: (1) A = Adj and B = that + Adj
or: (2) A = Man_x, and B = Man_x

and where: Man_x is a particular subclass of Manner Adverbials of the form:

$$Man=(Adv_a)Adj+ly$$

or its equivalent, where Adv_a is an Attributive Adverbial,
and where: Z + B + W is not Negative.

Reductions:

T1.

$$X\begin{bmatrix} as+Adj+Y+as \\ more+Adj+Y+than \end{bmatrix}Z+that+Adj+W \rightarrow X\begin{bmatrix} as+Adj+Y+as \\ more+Adj+Y+than \end{bmatrix}Z+Adj+W$$

T2. $$X\begin{bmatrix} as+Adj_1+Y+as \\ more+Adj_1+Y+than \end{bmatrix}Z+Adj_2+W \rightarrow X\begin{bmatrix} as+Adj_1+Y+as \\ more+Adj_1+Y+than \end{bmatrix}Z+W$$

where $Adj_1 = Adj_2$

T3. (Optional)

$$X+Vb_1\begin{bmatrix} as+Adj+Y+as \\ more+Adj+Y+than \end{bmatrix}Z\left(\begin{bmatrix} Ing \\ En \end{bmatrix}\right)Vb_2+W \rightarrow$$

$$X+Vb_1\begin{bmatrix} as+Adj+Y+as \\ more+Adj+Y+than \end{bmatrix}Z+W$$

where $Vb_1 = Vb_2$

T4. (Optional)

$$X\begin{bmatrix} have+En \\ M \end{bmatrix}be+Y\begin{bmatrix} as+Adj+Z+as \\ more+Adj+Z+than \end{bmatrix}W\begin{bmatrix} have+En \\ M \end{bmatrix}be+U \rightarrow$$

$$X\begin{bmatrix} have+En \\ M \end{bmatrix}be+Y\begin{bmatrix} as+Adj+Z+as \\ more+Adj+Z+than \end{bmatrix}W\begin{bmatrix} have \\ M \end{bmatrix}U$$

[18] The conventions for writing grammatical rules are those of Lees, *ibid.*

T5. (Optional)

$$X+M+\text{have(Y)}\begin{bmatrix}as+\text{Adj}+Z+as\\more+\text{Adj}+Z+than\end{bmatrix}W+M+have+U \rightarrow$$

$$X+M+have\text{(Y)}\begin{bmatrix}as+\text{Adj}+Z+as\\more+\text{Adj}+Z+than\end{bmatrix}W+M+U$$

T6. (Optional)

$$X+\text{Tns}\begin{bmatrix}M_1\\\text{Aux}_{b_1}\end{bmatrix}Y\begin{bmatrix}as+\text{Adj}+Z+as\\more+\text{Adj}+Z+than\end{bmatrix}\text{Nom(Prev)Tns}\begin{bmatrix}M_2\\\text{Aux}_{b_2}\end{bmatrix}Y+W$$

$$\rightarrow X+\text{Tns}\begin{bmatrix}M_1\\\text{Aux}_{b_1}\end{bmatrix}Y\begin{bmatrix}as+\text{Adj}+Z+as\\more+\text{Adj}+Z+than\end{bmatrix}\text{Nom}+W$$

where: Y is in the main clause of the Constituent-sentence.

In conclusion I should like to take note of several other features of the so-called Comparative constructions. Notice to begin with the following triple ambiguity:

(97) He had known more attractive women than Annie.[19]

which can be construed either as:

(98) He had known women more attractive than Annie.
(99) He had known more attractive women than Annie had known.
(100) He had known attractive women other than Annie.

Clearly **(99)** is an ordinary Comparative construction such as we have just analyzed. When **(97)** is construed like **(100)** it appears to contain a colloquial stylistic variant of the word *other*, as in:

(101) He had known more attractive women than just Annie.

But **(98)** can be construed only as a normal Post-Nominal Modifier construction from the Relative-Clause sentence:

(102) He had known women who were more attractive than Annie.[20]

What then is the original triply ambiguous sentence **(97)** when construed in this latter way?

When the nominal in question is singular, we have:

(103) He had known more attractive a woman than Annie.
(104) He had known a woman more attractive than Annie.

the latter being again a reduction from:

(105) He had known a woman who was more attractive than Annie.

[19] The example is taken from Jespersen, *op. cit.*, p. 397.
[20] Lees, *ibid.*, pp. 85–94.

Thus, we seem to have a special inversion of comparative post-nominal modifiers:

a book that good → a book better than that →
a better book than that
books that good → books better than that →
better books than that

Taking the "equative" form of the Comparative construction, we have:

a book that good → a book as good as that →
as good a book as that
books that good → books as good as that →
as good books as that

Thus, we see that this inversion operates in a slightly different manner in the two Comparatives; the so-called "comparative-degree" adjective follows the indefinite article, but the equative *as* + *Adj* precedes it.

There is also a special effect on the so-called *It*-Inversions, that is, stylistic variations of the form:

(106) That he goes is good. →
(107) It is good that he goes.[21]

When the Adj is comparativized first, then one can have two different orders:

(108) That he goes is better than that. →
(109) It is better than that that he goes.

(110) That he goes is better than that I go. →
(111) It is better that he goes than that I go.

Finally, I must comment briefly on the overall significance of the Comparative Transformations. The Comparative rule serves to expand the Attributive Adverbial into an infinite class of expressions in the same way that Nominalizations do the Nouns,[22] the Complement transformations the Verbs,[23] the Relative-Clause and Post-Nominal Modifier rules the Adjectives, and various subordinate-clause constructions in *if*, *when*, *although*, *whether*, *since*, etc. do the non-attributive adverbial.[24] This fact is what characterizes the major syntactic categories of the language, and is,

[21] Lees, *ibid.*, p. 94.
[22] *Ibid.*, Chapter III.
[23] *Ibid.*, p. 9, and references to Chomsky therein.
[24] I.e., rules which would expand, say, *then* in *He came then* to *when I left*, a transform of *I left*, yielding: *He came when I left*, etc.

in my view, a far more fundamental feature of major categories than is the fact that these usually also contain a large but finite lexicon of single-word members, as in the cases of Noun and Verb.[25] These transformations extend the finite multiplicity of certain grammatical categories to that of at least one denumerably infinite subset of sentences.

International Business Machines Research Center
P.O. Box 218
Yorktown Heights, New York

[25] Lees, "The Grammatical Basis," pages 9–10 and footnotes 19–21.

JOSEPH R. APPLEGATE

Phonological Rules of a Subdialect of English

In many cases, the attention of the linguist is given to the observation of language only as it is used in normal communication, i.e., among the majority of the normal adults in a community. The observation and description of linguistic anomalies or deviations are often left to the psychologist, the neurologist, or an investigator from some other field. While it is true that normal speech and communication are basic areas of linguistic research, it is also true that consideration of deviations from normal speech patterns is useful, and necessary, for a clear formulation of structural characterizations because of the insight it provides into the factors determining the development of those characteristics.[1]

Among the deviant forms of language is the speech of children. Usually the deviations found in children's speech are ascribed to the fact that the children have not acquired all of the patterns of adults' speech. A study of the speech of children, then, is essentially a study of the stages in the development of the language of the adult community. The development of speech is a slow process, and since the formation of language habits (native language patterns) cannot be duplicated under laboratory conditions with controlled variables, it has not been possible to design experiments to test all of the various hypotheses that have been proposed. The best that can be done is to study a large number of samples of speech at various stages of development and from these observations to reconstruct the process.[2]

This paper is a report of the investigation of the speech of three children. The subjects are brothers, who when the study began were four, five and

[1] One such study, "Kindersprache, Aphasie und allgemeine Lautgesetze," by R. Jakobson, *Uppsala Universitets Årsskrift* (1942) has proved extremely important in the development of a framework for the description of the phonological features of language.

[2] Studies such as those of W. F. Leopold, *Speech Development of a Bilingual Child* (1939–1949), Northwestern University Press, Evanston, Ill., 4 vols., furnish a great deal of information about the development of speech in individuals.

one-half, and eight and one-half years old. They are the sons of native-born American parents, whose speech is typical of the small New England community near Boston in which they live. Both parents were born in that area and attended public schools there. Their formal education ended with the completion of secondary school. Although there were no obvious deviations from normal speech patterns in the speech of the parents, the father reported a history of stuttering. His age at the onset of stuttering could not be determined exactly, but his mother confirmed his statement that it began after he had started attending school. The stuttering continued throughout childhood and adolescence and disappeared during the period of military service during World War II. Later, the father reported that his stuttering might still return in situations involving some emotional stress. As an example of such a situation, he mentioned applying for a job, grouping this with similar situations in which he ". . . wanted to make a good impression." He described his stuttering as ". . . difficulty with the breath sounds, like 'h' and 'f' and maybe 's'." Usually, however, there was no stuttering when he spoke, and the syntactic and morphological patterns in his speech, as well as in that of his wife, were those of the local dialect.

The parents realized that there were deviations from the local dialect in the speech of all three of the children; but the oldest son had begun to stutter, and they hesitated to correct the speech of the two younger children because they feared that they too might develop the same difficulty. This fear stemmed from the fact that they recognized in the younger children's speech the same deviations that had once been characteristic of the oldest son's speech. (It is possible that there was some relationship between attempts to correct the child's speech patterns and the development of stuttering, but it is not the purpose of this paper to explore that aspect of the case.) Attention will be concentrated on the description of the younger children's speech in which the deviations that occurred were remarkably similar. In fact, the language of these two children can be considered a subdialect of English. It was the only language spoken by them, and they used it in communicating with the adults of the community, who spoke English, as well as in communicating with each other and with their playmates.

It is this last feature that makes the case especially interesting, for usually the speech of a child is recognized as a deviant form of the language of adults in the community, and the child is compelled, in a number of ways, to change his speech until it conforms to the patterns of the adults' language. This is due to the fact that the speech patterns of the people around him are quite different from his own, and he is, therefore, always in the position of communicating with people who speak a different dialect.

Even his playmates, because of environmental influences, speak languages that are quite different from his own, although their speech may also contain patterns that deviate from those of the adult community. This was not true of the children in this case study; they were able to communicate freely with each other, with their playmates, and usually with the adults with whom they had frequent contact. The oldest brother, whose speech was much more similar to the language of the adult community, understood the speech of his brothers and served as a translator for them in situations in which they could not communicate effectively with adults. (This includes situations in which the parent, because of fatigue or impatience, failed to understand the younger children.) This fact may be significant in accounting for the preservation of the speech habits of the younger children. Pressure to make themselves understood—that is, pressure to alter deviant speech habits—was minimized because their brother was willing to serve as a translator for them if communication broke down.

Although the syntactic and morphological patterns of this subdialect were very much like those of standard American English, some deviations were readily apparent. One of these, in syntax, was the omission of subordinating conjunctions. This occurred not only in indirect statements or relative clauses, but in other types of subordinate clauses as well. In morphology, the most obvious deviation was the partial identity of past-tense and third-person singular (present tense) verb affixes with the plural and possessive affixes for nouns. Each of these included among its alternants the following phonetic forms: [-t], [-d], and [-id]. Thus the form [wakt] was heard in contexts that served to identify it clearly as both past-tense and third-person singular, present tense of the verb 'walk', and in one case as a plural noun (for 'sidewalks'). In addition to these suffixes, a zero suffix with change of stem vowel was noted among the alternants of the past-tense morpheme as well as among the alternants of the plural morpheme in such forms as [ræn] from 'run' and [men] from 'man.' The zero suffix without change of stem vowel was an allomorph of the third person singular morpheme when it occurred with the modal auxiliaries, and it is this fact which serves to separate these three from the possessive morpheme. There was no zero allomorph for the possessive. The patterns in which the zero alternant was used for plural nouns and for verbs were rather unstable, however, and there seemed to be a tendency to have zero alternate freely with one of the other affixes without change of stem vowel, so that the forms [rənd], [mænd], and [kænd] occurred for 'ran', 'men', and 'can'. (This instability is not unusual in the speech of children and may be interpreted as a tendency to regularize patterns because of the limited number of alternants that have been learned for each morpheme.)

The feature that served to separate the past tense morpheme from the present tense and noun plural morphemes was the inclusion of the forms [-ʔ] and [-iʔ] among the allomorphs of the past tense.[3] The past of 'do' was [diʔ], while the present was always [dəd]. The past tense of 'talk' was [takiʔ], but the present tense form was [takt].

The deviant morpheme alternants may be included in the description simply as deviant forms that are characteristic of the dialect, and a correlation of these forms with those of standard speech might be established. It was suggested that the children might have adopted a single affix for all verb forms, but this solution can be rejected immediately because of the existence of [-ʔ] and [-iʔ] as alternants of the past-tense morpheme but not of the present-tense morpheme. Furthermore, there was no evidence of syntactic confusion of nouns and verbs, a necessary condition for explaining the identity of noun plural and third-person singular affixes in terms of the proposed solution. Since the deviations are completely regular, not sporadic, phenomena, it seems best to treat them as part of the autonomous system that is the children's dialect.

Further examination of the data shows that the distribution of the forms [ʔ] and [-iʔ] as alternants of the past-tense morpheme is highly restricted. These alternants occur only in sequences in which there is a preceding stop such as [diʔ] 'did', [dayʔ] 'died' and [teykiʔ] a form which alternated with [tuk] 'took'. A second restriction was found in the occurrence of the glottal stop only in medial or final position. These facts suggest that the occurrence of [-ʔ] may be due to some phonological feature of the children's dialect. In other utterances the glottal stop occurred under the same conditions, but in words that were clearly not verb forms. The children called their father [dæʔiy]. A playmate whose name was Bobby was called [baʔiy]. The word 'paper' was [peyʔər]. The glottal stop never occurred initially, but only in medial or final position: [pəʔiy], [keyʔ], [kiʔ], [dæʔiy], [beyʔiy].

A tentative solution postulating the occurrence of a glottal stop in medial or final position in any sequence containing a preceding stop in initial or medial position must be rejected because of the occurrence of such forms as [pæt], [tæg], [bit], [biyt], and [pet]. The glottal stop is used in the past tense, however, only if the word contains a preceding [t] or [d]. It is possible, then, to assume that the second, third, or *n*th occurrence of a stop within a word requires the replacement of that stop by a glottal stop.

[3] The term 'present tense morpheme' here and in the following sections of the paper refers only to the third-person singular form of the present tense. The third-person plural, as well as the first- and second-person forms in this dialect corresponded to those of standard American English.

This is supported by the examples of other words cited above; all of them should contain two identical stops. This solution is only partially correct, however, for words in which identical stops occur do exist in the dialect. Such forms as [takt], [teykt], the present tense forms of 'talk' and 'take', as well as [dagd], the plural of 'dog', illustrate this.

It must be assumed that there are two series of stops in the children's dialect. Stops from the first series /C′/ alternate with glottal stop if preceded by an identical /C′/. Stops from the second series /C″/ remain unchanged. The past-tense morpheme can then be written as /t′/ alternating with /d′/, and the present-tense, noun plural and possessive morphemes can be written as /t″/ alternating with /d″/. In the phonology, a rule can be included to describe the change. For the proper application of such a rule, it is necessary that both stops be identical. It is possible to specify the features of the stops in terms of the distinctive feature framework developed by Jakobson and Halle.[4] The rule stated in distinctive feature terminology has the following form:

$$\#(X)+\begin{bmatrix}+\text{consonantal}\\-\text{vocalic}\\-\text{continuant}\\ x\end{bmatrix}+X+\begin{bmatrix}+\text{consonantal}\\-\text{vocalic}\\-\text{continuant}\\ x\end{bmatrix}+(X)+\#\Rightarrow$$

$$\#(X)+\begin{bmatrix}+\text{consonantal}\\-\text{vocalic}\\-\text{continuant}\\ x\end{bmatrix}+X+\begin{bmatrix}-\text{consonantal}\\-\text{vocalic}\\-\text{continuant}\end{bmatrix}+(X)+\#$$

In this rule '#' represents word boundary.[5] 'X' is any phoneme or sequence of phonemes not having the same features as those listed in

[4] R. Jakobson and M. Halle, *Fundamentals of Language* (Mouton and Co., 'S-Gravenhage, 1956). I am especially indebted to Prof. Halle for his helpful comments during the preparation of this paper.

[5] It is possible to set the word boundary as a limit for the change because of the occurrence in the corpus of sentences such as (#dae?iy#di?#it#) in which the first stop of each of the first two words is clearly articulated as (d), but in which the second stop of each of these words is replaced by (?). This indicates that the change does not continue beyond word boundaries. If it did, the initial (d) of the second word would be replaced by (?) also.

brackets. 'x' represents any permissible set of features, but these must be identical in each of the bracketed sets. Parentheses indicate that the element is optionally present, so that the first stop may be in either initial or medial position; the second is either medial or final. Application of the rule results in the change of one distinctive feature in the second bracketed set, i.e., the segment becomes non-consonantal. The second and third features for that segment remain unchanged. Features after the third may be deleted, for the only phonetic realization of a nonconsonantal, nonvocalic, noncontinuant is the glottal stop. The problem of distinguishing /C′/ from /C″/ remains, however. It is necessary to make such a distinction so that the second stop remains unchanged if it is /C″/ or if it is /C′/ preceded by /C″/. All stops of /C′/ must have a feature not found in /C″/ or vice versa if this distinction is to be made.

At this point, reference to the language of the community is helpful. In adult speech, the form of the plural morpheme, the possessive, and the third-person singular verb morpheme is [s] alternating with [z] in those places where the children have [t] and [d]. Further examination of the children's speech shows that all fricatives are missing. /f/ is articulated as /p/, /v/ is /b/, /s/ is /t/ and /z/ is /d/. The interdental fricatives are also articulated as stops, and /c/ and /j/ are /t/ and /d/ respectively. The two series of stops in the children's language are then related in the same way that stops and fricatives are related in adult speech. A rule of the following form, again in terms of distinctive features can be used to describe this phenomenon:

$$\begin{bmatrix} +\text{consonant} \\ -\text{vocalic} \\ +\text{continuant} \\ \text{x} \end{bmatrix} \rightarrow \begin{bmatrix} +\text{consonant} \\ -\text{vocalic} \\ -\text{continuant} \\ \text{x} \end{bmatrix}$$

The application of this rule results in the change of all fricatives, i.e., continuants, to noncontinuants, or stops. The other features of the segment, represented in the formula by 'x', remain unchanged so that voicing, for example, is not affected. This accounts for the fact that /f/ is articulated as [p] and never as [b].[6]

These two rules are sufficient to describe the phonetic deviations in the speech of the children, but an apparent contradiction arises. Application

[6] It is interesting that the consonants affected by the special rule that converts fricatives to stops are those that are most affected by stuttering in the speech of the oldest child. Apparently, this is also true of the father's speech. This suggests that there is a relationship between the attempts made to correct deviant patterns and the later development of anomalies in speech.

of the second rule converts a continuant to a noncontinuant, which may have features identical with those of a preceding noncontinuant. If the first rule is then applied, the second consonant will become a glottal stop. This cannot be permitted because it does not conform to the facts observed. The contradiction can be eliminated by imposing the requirement that the rules be applied in a fixed order: the first rule, which converts a duplicated stop to a glottal stop, must always be applied before the second, which converts fricatives to stops. Then only those stops indicated above as /C′/ will be changed to glottal stops; stops in the series /C″/ are produced only after this change has taken place. This accounts for the contrasting forms [takiʔ] 'talked' and [takt] 'talks', or [tuwʔ] 'toot' and [tuwt] 'suit'. After the application of these two rules, subsequent phonetic rules are applied to both series of stops alike so that each of the remaining stops of C′ has allophones that are identical with those of the corresponding stop of C″. (The voiceless stops are aspirated in initial prevocalic position, for example.)

The place of these rules in the grammar of the children's dialect can be specified more exactly by using these facts. The fact that [-ʔ] occurs as an alternant of the past tense morpheme shows that /t/ or /d/ must be added to the verb stem before the application of the phonological rule that converts duplicated stops to glottal stops. If this were not true, the glottal stop would never occur in such sequences. The fact that those stops of /C′/ that do not become glottal stops have allophones like those of the corresponding stops of /C″/ shows that the rules affecting these consonants must be applied before the rules that determine the phonetic features of positional variants. Thus, the rules that are peculiar to the children's dialect are not mere appendages to the grammar but an integral part of the set of rules for generating sentences in that dialect. They are ordered not only in relation to each other but also in relation to the other rules of the grammar. It is this feature of the description that clarifies the relationship between the grammar of the children's dialect and the grammar of the dialect of the adult community. The two are almost exactly the same: the syntactic rules are practically identical;[7] the morphology is largely the same for the two languages; and the major difference is found in the inclusion of these two phonological rules in the grammar of the children's dialect, but not in that of the adults' dialect. A description in which a different inventory of phonemes was postulated or in which different rules for the distribution of phonemes were included would tend to obscure the structural similarity of the two dialects, for it would necessitate the description of different sets of alternants for the same morpheme in the two dialects. The description

[7] An exception is found in the case of subordinating conjunctions, as mentioned above.

presented in this paper is more satisfactory, for it highlights the similarity of two dialects that are known to be related. Furthermore, it localizes the point of difference that gives the children's speech its distinctive character.

This description also provides more insight into the development of language. On the basis of our observations, we may say that the children's speech does not represent a random attempt to imitate the language of the adult community. Instead, it is clearly an autonomous system with well developed rules.*

Department of Near Eastern Languages
University of California
Los Angeles 24, California

* The work presented in this paper was completed as part of a study of translation supported in part by the National Science Foundation while the author was a staff member of the Research Laboratory of Electronics of the Massachusetts Institute of Technology. Work in the Research Laboratory of Electronics is supported in part by the U.S. Army (Signal Corps), the U.S. Air Force (Office of Scientific Research, Air Research and Development Command), and the U.S. Navy (Office of Naval Research).

IVÁN FÓNAGY

Communication in Poetry

1. Though we may be unable to predict the subjects discussed at a given gathering, we can fairly accurately determine in advance how often individual sounds will occur in the course of a conversation. We can tell the printer approximately how many *a*'s, *f*'s and *g*'s he will need for a 32-page short story. All we need to know for this is the relative frequency of occurrence of the sounds of the language concerned, as established by statistical analysis.

Yet, the actual truth always holds a certain element of surprise. It may turn out that the author of the short story or the guests at our party use some sounds more or less often than we would have expected on the basis of previous experience. (Nor will accuracy of our expectation increase significantly if we know what topics will come up in the course of the conversation, or whether the short story has a happy or a tragic end.) The very statistical data on which we base our prediction stemming from the analysis of different sets of materials show a certain divergence. Thus, the relative frequency of /l/ in Hungarian is 5.9% according to T. Tarnóczy (who analyzed Endre Ady's poems), 4.98% according to E. Vértes (who examined the prose of Péter Veres), and 6.27% according to the records of a parliamentary stenographer.[1]

2. It is somewhat easier to make predictions if we are dealing with poetry. If we know the topic and basic mood of a given poem, we can predict in what direction and approximately in what measure the relative frequency of certain sounds is likely to deviate from an index number based on statistics of the standard language. M. M. Macdermott, for instance, through a statistical analysis of English poems, found that dark vowels are more frequent in lines referring to dark colours, mystic obcurity, or slow and heavy movement, or depicting hatred and struggle.[2]

Paul Guiraud analyzed the distribution of sounds in Valéry's poems. He found, among other things, that in *Les grenades*, deliberately done in crude

[1] E. Vértes, "Phonetischer Aufbau der ungarischen Sprache," *Acta Linguistica Academiae Scientiarum Hungaricae*, III (1953), 136.

[2] *Vowel Sounds in Poetry, Their Music and Tone-colour*. Psyche Monographs, No. 13, London, 1940, 89–92.

colours that suggest a cheerful aggressiveness (as much as this is possible in a still life of pomegranates), the "sharp" and "clear" vowel-tones are dominant, and among the consonants, besides /r/, generally the velar and labial stops occur with greater than average frequency. In *Dormeuse*, Valéry describes a sleeping woman with tenderness and sensuality (*langueur* is the key-word of the poem), and here the back vowels prevail and the nasal vowels and consonants are relatively frequent, while the velar and dental stops lose their importance.[3]

Analysis of the sounds in six especially aggressive and six especially tender poems of the great and impulsive Hungarian lyric poet Sándor Petőfi reveals the following distribution.[4] (See Table 1.)

The majority of the sounds occur with about the same relative frequency in both groups. All the more striking is the fact that the frequency of certain sounds shows a significant difference in the two groups. The phonemes /l/, /m/, and /n/ are definitely more frequent in tender-toned poems, whereas /k/, /t/ and /r/ predominate in those with aggressive tone. For some reason precisely these sounds seem to be the most significantly correlated with aggression—either positively or negatively. The back vowels are also generally more frequent in the aggressive poems. But the tense and high /i/, which according to Grammont may express anger,[5] also occurs frequently in the aggressive poems.

Of course, it may be supposed that the relatively greater frequency of /t/, /k/, and /r/ in words suited to the expression of aggressive emotions reflects a peculiarity of the Hungarian language. The question poses itself whether the "sound spectrum" shows a similar modification in a count based on French or German poems.

A comparison of some of Hugo's and Verlaine's angry and tender poems[6] (see Table 2) reveals that the sounds /t/ and /k/, and to a lesser extent /g/ and /r/, are dominant in Hugo's fiercer poems. The tender

[3] *Langage et versification d'après l'œuvre de Paul Valéry*. Paris, 1953, pages 89f., 147f.

[4] The aggressive poems: *Egy goromba tábornokhoz* (To a Rude General), *Hallod-e szív?* (Hearest thou, Heart?), *Nyakravaló* (Hempen Collar), *Akasszátok fel a királyokat!* (Let's Hang the Kings!), *Feltámadott a tenger* (The Sea has Risen), *Austria*. The tender poems: *Messze vándoroltam* (I've Roamed Afar), *Borús, ködös, őszi idő* (Dark and Foggy Autumn Weather), *Alkony* (Dusk), *Minek nevezzelek?* (What Shall I Call Thee?), *Itt van az ősz* (Autumn has Come), *Szülőföldem* (My Birthplace).

[5] M. Grammont, *Le vers français*. Paris, 1947, p. 250.

[6] The aggressive poems: V. Hugo: *L'homme a ri, Napoléon III, Au peuple;* P. Verlaine: *Monsieur le docteur Grand, A un magistrat de boue*. The tender poems: Hugo: *Quand l'enfant paraît, Trois ans après, A Granville*; Verlaine: *Ecoutez! Parabole, Soleils couchants, Chanson d'automne*. First I determined the proportion of the different sounds for each of the two categories for each poet and then took the mean of these figures.

TABLE 1

		Petőfi's poems *Tender*	Petőfi's poems *Agressive*	Standard Hungarian
STOPS VOICELESS	k	7.74	10.70	9.14
	t	10.70	13.62	13.30
	p	1.35	1.32	1.45
	Total	*19.79*	*25.64*	*23.89*
STOPS VOICED	g	4.11	3.57	4.32
	d	4.95	4.38	3.67
	b	2.92	2.89	3.37
	Total	*11.98*	*10.84*	*11.36*
AFFRICATES	c	0.22	0.45	0.38
	č	1.42	0.86	0.96
	Total	*1.64*	*1.31*	*1.34*
FRICATIVES VOICELESS	s	3.62	3.43	3.27
	š	5.02	4.13	6.54
	f	1.31	2.04	1.58
	h	3.24	3.33	3.03
	Total	*13.19*	*12.93*	*14.42*
FRICATIVES VOICED	z	2.36	2.89	3.99
	ž	0.10	0.43	0.13
	v	3.92	3.49	3.37
	Total	*6.88*	*6.81*	*7.49*
PALATALS	t′	0.95	0.54	0.17
	d′	2.51	2.45	2.60
	ń	0.95	1.40	1.27
	j	3.97	2.70	3.01
	Total	*8.38*	*7.09*	*7.05*
LIQUIDS	r	5.77	7.60	7.25
	l	12.51	10.58	10.05
NASALS	[ŋ]	0.85	0.44	—
	n	9.42	8.02	9.74
	m	9.62	7.79	7.09
	Total	*19.89*	*16.25*	*16.83*

	Petőfi's poems *Tender*	Petőfi's poems *Agressive*	Standard Hungarian	
i	0.48	0.89		LIGHT VOWELS
ī	8.75	8.71	11.26	
ü	0.73	0.54		
ū	1.17	0.97	1.79	
ë	7.16	6.99	8.45	
ɛ	31.47	23.48	26.04	
ő	2.30	2.13		
ö	2.85	3.36	4.90	
ä	8.36	8.55	8.48	
Total	*63.27*	*55.62*	*60.92*	
a	20.98	25.81	23.57	DARK VOWELS
ō	3.03	2.64		
o	9.52	12.15	12.14	
u	0.92	1.15		
ū	2.28	2.63	3.17	
Total	*36.73*	*44.38*	*38.88*	

TABLE 1. PERCENTAGE OF CONSONANTS AND VOWELS IN ESPECIALLY TENDER AND AGRESSIVE POEMS BY PETŐFI AS COMPARED TO STANDARD HUNGARIAN. (THE PALATAL NASAL /ń/ IS HERE COUNTED AS A PALATAL RATHER THAN A NASAL. THE VELAR NASAL [ŋ] IS AN ALLOPHONE OF /n/. "DARK" VOWELS ARE CORRELATED WITH BACK-ROUNDEDNESS. THE GRAPHIC SEQUENCE *ly* HAS BEEN COUNTED WITH /j/. PHONETICALLY, /a/ IS SLIGHTLY ROUNDED.)

TABLE 2

	Tender Poems			Aggressive Poems		
	Hugo	Verlaine	*Mean value*	Hugo	Verlaine	*Mean value*
k	6.83	7.65	7.24	7.69	9.56	8.63
t	7.47	6.52	7.00	14.58	14.18	14.38
p	5.53	3.78	4.65	5.09	5.13	5.11
Total	*19.83*	*17.95*	*18.89*	*27.36*	*28.87*	*28.12*
g	0.96	1.63	1.30	1.35	1.07	1.21
d	7.09	7.93	7.51	5.76	10.11	7.94
b	3.68	2.80	3.24	3.56	2.17	2.86
Total	*11.73*	*12.36*	*12.05*	*10.67*	*13.35*	*12.01*
s	8.50	10.08	9.29	7.81	10.15	8.98
š	1.56	2.17	1.87	1.47	—	0.74
f	3.78	3.40	3.59	3.49	4.19	3.84
Total	*13.84*	*15.65*	*14.75*	*12.77*	*14.34*	*13.56*
z	3.52	1.05	2.29	2.10	1.19	1.64
ž	2.68	2.21	2.44	1.91	0.26	1.09
v	5.17	5.64	5.41	3.42	4.56	3.99
Total	*11.37*	*8.90*	*10.14*	*7.43*	*6.01*	*6.72*
ɲ	0.05	0.27	0.16	0.06	0.51	0.29
j	2.35	3.85	3.10	2.05	1.07	1.56
ʎ	0.84	0.73	0.79	0.70	0.73	0.71
Total	*3.24*	*4.85*	*4.05*	*2.81*	*2.31*	*2.56*
r	16.03	13.13	14.58	16.95	15.39	16.17
l	13.70	14.94	14.32	13.33	8.48	10.91
n	4.21	5.10	4.66	3.38	4.56	3.97
m	5.94	7.24	6.59	5.39	6.83	6.11
Total	*10.15*	*12.34*	*11.25*	*8.77*	*11.39*	*10.08*

TABLE 2. FREQUENCY (IN PERCENTAGES) OF CONSONANTS IN ESPECIALLY TENDER AND AGGRESSIVE POEMS OF VICTOR HUGO AND PAUL VERLAINE.

poems are characterized chiefly by the relative scarcity of "piercing" sounds. In Verlaine's aggressive poems there is a striking predominance of /t/'s, but there is a higher than usual frequency of /k/'s (as a matter of fact of all stops) and /r/'s, as well. In the gentler verses /l/ ranks high (14.94% as compared to 8.48% in the aggressive poems), together with /j/, /z/, /ž/, /v/, and /n/, and in a smaller measure also /m/. The vowels seem to reflect the change in mood to a lesser extent. The vowel /ə/, articulated "passively," seems to be negatively correlated with aggressiveness; /a/ and /ü/ occur more often in angry poems.

The differences under discussion are even more evident when cycles of aggressive poems are compared with cycles of idyllic ones, e.g. the third

volume of V. Hugo's *Châtiments* with the first three cycles of *L'art d'être grand père* and with the twelfth and seventeenth cycles about the sleeping Jeanne, or the first 50 poems of Verlaine's *Invectives* with his *Bonne chanson*. Analysis of Petőfi's lyric poems from 1847 until his death in 1849, separating the revolutionary or angry poems from the idyllic poems, reveals smaller deviations from the mean, but here again /t/ and /k/ and /r/ are dominant in the aggressive poems, and there is a negative correlation beween anger (revolutionary or personal) and the "soft" /m/ and /l/. In Hugo's and Verlaine's poems /r/ turned out to be neutral, which is probably due to the "softer" (velar or uvular) pronunciation of the modern French /r/.[7]

Nevertheless, the fact that the correlation between the frequency of /r/ and aggressiveness is not limited to Hungarian poetry is indicated by statistics of two cycles of poetry by Rückert, the first "bouquet" (*Erster Strauss*) of the idyllic *Liebesfrühling*-cycle,[8] and the first 32 poems of *Geharnischte Sonette*.[9] In these militant "armoured" sonnets the /t/'s, /r/'s, and to a lesser extent the /k/'s are dominant. In the *Liebesfrühling* there is a prevalence of /l/'s and /m/'s (see Table 3 and Figure 1).

TABLE 3

	Tender Poems					Aggressive Poems				
	Hugo	Verlaine	Petőfi	Rückert	*Mean value*	Hugo	Verlaine	Petőfi	Rückert	*Mean value*
k	2.85	3.03	5.03	1.80	3.18	3.22	3.88	7.01	1.93	4.01
t	3.97	4.14	7.04	5.15	5.07	5.19	5.22	9.01	7.26	6.67
r	9.38	8.30	3.72	6.52	6.98	9.10	8.60	4.91	7.75	7.59
l	7.91	6.61	6.25	5.29	6.50	6.81	6.35	6.10	3.73	5.75
m	3.16	3.29	5.72	3.48	3.91	2.18	3.56	4.80	2.95	3.37

TABLE 3. FREQUENCY OF FIVE CONSONANTS IN TENDER AND AGGRESSIVE POEMS BY VICTOR HUGO, PAUL VERLAINE, SÁNDOR PETŐFI, AND FRIEDRICH RÜCKERT. (IN PERCENTAGES.)

3. On the semantic level, too, we would expect the elements to be more predictable in poetry; one would think that it is easier to guess the next word in a poem than in every-day speech or in a newspaper article.

But the facts show the very opposite. Experiments on twentieth-century Hungarian poems, modern newspaper articles, and the texts of recorded

[7] In the opinion of French estheticians the velar or uvular /r/ ("*r*-grasseyé") is not hard. Cf. A. Spire, *Plaisir poétique et plaisir musculaire*. Paris, 1949; pages 195, 234, 364.

[8] Friedrich Rückerts *Werke*, Stuttgart: Cotta, 1895; 9–30. The cycle contains a total of 16,378 sounds.

[9] Friedrich Rückerts *Werke*, I, 3–16. The 32 poems contain 14,207 sounds.

conversations show that the sequence of words in the sentence is the least predictable in (modern) *poetry* (cf. Figure 2).

In the experiments described below, sixty-one per cent of the sounds in the poems of Endre Ady had to be provided by the tester, whereas in the case of a newspaper article it was sufficient to give 33% of the sounds to enable the subjects to guess the remaining sounds (i.e. the words). Thus

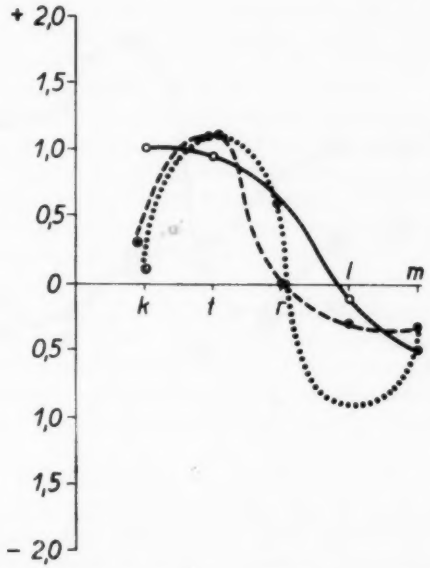


FIGURE 1. Relative frequency of five consonants in aggressive Hungarian (——), French (– – –), and German (· · · ·) poems, as compared to the mean values of these consonants in the work of each poet considered.

semantic predictability varies considerably with genre. (It should be remarked here that I deliberately chose the most run-of-the-mill type of articles.) In a recorded conversation of two young girls it took the experimental subjects 5549 guesses to ascertain the 13,411 missing sounds. This means that only 29.3% of the conversation was of informational value.

Our method[10] in determing the semantic redundancy of the sentences was as follows: Each of twenty subjects was told that he or she was to guess a

[10] I am greatly indebted to my associate, Dr. Klára Magdics, for her assistance in carrying out the experiments.

few sentences sound by sound. The tester supplied the first sound of the passage, and the subject was to guess the following sound, word, or even sentence. At each point where the subject was unable to guess correctly (and at those points at which the elements in a correct guess ended) the

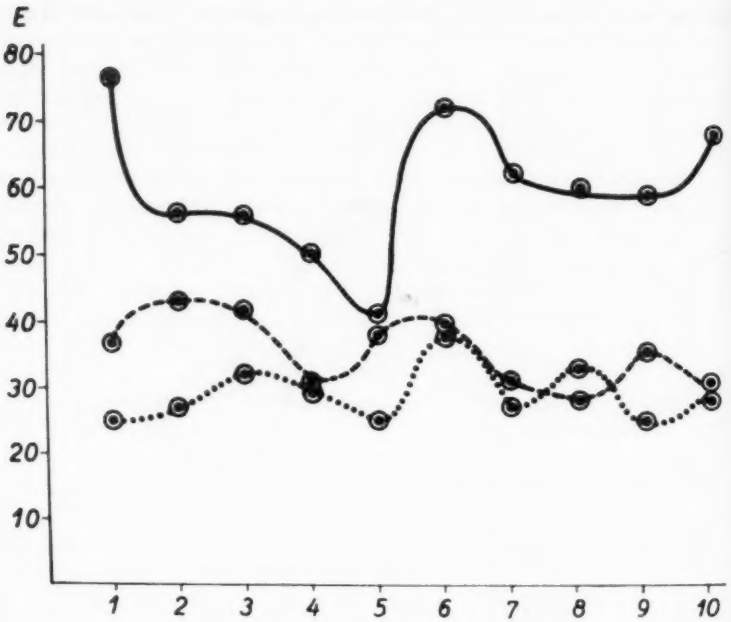


FIGURE 2. Informative value of ten lines of a modern Hungarian poem (——), of ten sentences in a run-of-the-mill type of article (– – –), and in a recorded conversation of two young girls (· · · ·). Vertical axis: percentage of sounds that could not be identified in each sentence. Horizontal axis: the lines of poetry or sentences of prose, in temporal sequence.

tester would supply the next sound. At the end of the passage the percentages of guessed and supplied sounds could be determined and comparisons made with other passages.

To determine the transition probabilities of phonemes a less subjective method is employed; that is, the probability that an /l/ will follow a /k/ is determined by the analysis of texts rather than by an appeal to the expectations of the subjects of the experiment. But this method is hardly applicable when we are dealing with transitions between words, not sounds, for the

number of possible word combinations is vastly larger. In order to determine the linguistic norms of transition probability for words our corpus would have to be close to infinite, and until electronic computers can perform such operations for us we must rely on the human brain. Memory, of course, plays a most important role: The more frequent the linkage of two words is (in speech or writing), the more easily the second of the two is associated with the first when only the first one is known. Hence we postulate that subjective expectation corresponds to the objective transition probability of individual words. (A similar method was followed by Shannon and after him by others.[11] Our briefer procedure did not make it possible to determine entropy as defined by Shannon.)

However, we may pose the question whether readers familiar with the work of the poet (in this case Ady) would not have done better. Possibly the entropy of the poem examined is not so high *objectively* after all and the fault lies with the subjects of the experiment for failing to recognize the inherent syntactic and semantic rules of poetic language. There is no doubt that the quality of being different is also governed by rules, that divergence from the usual (or the unusual) turns used by other poets may become a rule. It is the great poets whose language is predictable and redundant in a certain sense. This is evident from the fact that their distinctive styles are easier to imitate and caricature than those of lesser poets. While the critic and reviewer evaluates an author on the basis of his predictable style and redundant forms, the philologist is able to identify him by these same characteristics.

Nevertheless, we can definitely say that the poem by Ady used in the experiment is less redundant than the run-of-the-mill newspaper article or conversation. Even if we should regard divergence from the usual as obligatory in poetry (in fact it is not obligatory: it cannot be determined in advance that the poet will reject the usual terms and combinations of expression), we would have to guess just which expressions he will select out of an infinite pool of unusual and unexpected turns. Predictability is much lower than it is if we can choose from a finite number of fairly stereotyped word combinations. Our experiments confirm that the laws determining the individual language of poets, the laws of unusual terms and combinations, are much more complicated than those of every-day language. Just as we preserved the anonymity of the journalist and that of the conversing girls, we did not disclose the name of the poet. Nevertheless,

[11] C. E. Shannon, "Prediction and Entropy of Printed English," *The Bell System Technical Journal*, XXX (1951); Frieda Goldmann Eisler, "Speech Production and the Predictability of Words in Contexts," *The Quarterly Journal of Experimental Psychology*, X (1958).

sooner or later, the majority of the experimental subjects recognized Ady. The individual linguistic idiosyncrasies, the predictable pat expressions of the journalist or the chatting girls could, of course, not have been known to the subjects. They were in a more favourable situation in the case of the poem, and of course, several of them had at some time heard or read this relatively little-known poem by Ady, though probably not very attentively. But despite all this, the unexpectedness of the poem was considerably greater.

In every-day language, strong threads of association link words together. These are the threads in which we get more or less deeply entangled when we translate into a foreign language. Charles Bally observes[12] that several attributes may modify the word *chaleur*. *Chaleur* may be *suffocante*, *accablante*, *tropicale*, *torride*, *sénégalienne*, etc.; whereas *froid* may be *épouvantable*, *mordant*, *cuisant*, *pénétrant*, *perçant*, *âpre*, etc. (In English *heat* is *oppressive*, *fierce*, *infernal*, *glowing*, etc.; while *cold* is *cutting*, *biting*, *pinching*, *piercing*, *shivering*, and so on.) This multitude of possible attributes may give one the illusion that it is possible to attach any rational modifier to *chaleur* and *froid*. In fact, this is not so. *Froid raidissant* (*froid qui raidit les membres*), a combination often found in Hungarian for instance, is rational, corresponds to "benumbing cold" in English, but it simply is not used. The same is true of other syntagms, such as the relationship of predicate and adverbial complement. The words *grièvement* and *gravement* are close synonyms; nevertheless, only a wounded person may be *grièvement blessé*, whereas someone who is ill is *gravement malade* (Bally, ibid., p. 70).

Similarly predictable and conventional are the similes of every-day speech. In German one drinks like a brush-maker (*Bürstenbinder*), in French like a fireman (*boire comme un pompier*), like a Pole (*comme un polonais*), or like a sponge (*comme une éponge*). In an English translation, however, we would unhesitatingly substitute fish for either the brush-maker, the Pole or the sponge, for that is the proper simile in English. *Bürstenbinder*, *pompier* and *fish* are so closely associated with the verbs *trinken*, *boire* and *drink* that the link is made by bypassing the real brush-maker, fireman, or fish.

The predictability of the conventional epithet is very great. In another experiment the attribute was left blank in a sentence taken at random from a newspaper article. Nine out of ten subjects guessed the attribute correctly even before the number of syllables in the attribute was given, while the tenth guessed it after the first sound was given.[13] In the Hungarian sentence

[12] *Traité de stylistique française*$_2$. Heidelberg–Paris, n.d., I–II.

[13] The experimental procedure was as follows: First the subjects were to guess the

FORRÓ *testvéri szeretetünket küldjük a baráti román népnek* 'We send our ARDENT fraternal love to the Rumanian friend-nation' six out of ten subjects guessed the underscored attribute that was left blank, one subject guessed it after the number of syllables was given, and three after the initial sound was given.

When the same adjective was left out of a prose version of a line of poetry by Árpád Tóth (*Forró hír repül feléjük a táviródróton át* 'Ardent news flows to them over the telegraph wire') three out of five subjects guessed right when they made their sixth guess (after the number of syllables and the first four sounds *forr-*). After the first two sounds *fo . . .* had been given the adjective *forró* occurred to several subjects, but they rejected it as absurd and, instead, guessed *fontos* 'important.'

In the syntagm *súlyos tévedés* 'serious error, grave mistake' ten subjects guessed the attribute correctly after the number of syllables or after the first sound had been given. The result was very different, however, when the syntagm *szeretõ tévedés* 'loving error' was taken from Attila József's *Alkalmi vers* 'Occasional Poem'. In this case five out of ten subjects guessed the adjective correctly only after five sounds had been given, and three subjects were not able to realize even after the fifth (i.e., last) sound had been given that they had the word. Likewise when the attribute in the syntagm *hûs zene* 'cool music' taken from Árpád Tóth's poem *Orgona* 'Lilacs' had to be guessed, several subjects insisted on a different attribute, namely *hûséges* 'faithful', even after the last letter had been given, and then gave up in despair. When in the combination *fanyar szobát* 'tart (or acrid) room' from the same poem the noun and the first two letters of the attribute had been given (*fa . . . szobát*), a great variety of suggestions were made (*fagyos* 'icy,' *fapadlós* 'wooden-floored,' *fakó* 'wan, sallow,' *faburkolatos* 'wood-panelled') but no one suggested the actual attribute used in the poem until the third sound (/ń/) had been supplied. Most subjects hesitated even after that, although, besides the word *fanyar* 'acrid' and its derivatives, the only other adjectives beginning with the graphic sequence *fany . . .*, are *fanyelû* 'with a wooden stem,' *fanyesõ* 'wood-clipping,' and *fanyomatos* 'printed as a boodblock'.

If, in Hungarian, a sentence begins with

Úgy sír mint. . . 'He cries like. . .'

our fancy races ahead, automatically continuing with *záporesõ* 'shower' (as in English one howls like a baby). But in Attila József's poem *Egy kisgyer-*

attribute left blank only on the basis of the context of the sentence. Then the person conducting the experiment divulged the number of syllables in the attribute, after that its first sound, then the second sound, and so on.

mek sír 'A Child is Crying,' the association leaves the trodden path with a sharp and unexpected twist:

> . . . as lime cries in water.
> He cries as. . .

Once again we are disappointed in our anticipation:

> . . . as water cries under the lid.
> He cries like dead wood consumed by fire.
> He cries like the conveyor belt in the shop. . .

The predictability of similes is likewise minimal in the poems of Eliot, Trakl, Rilke, or Apollinaire.

> Das Fieber donnert. Ihre Eingeweide
> *Brennen wie Berge. . .*
> (Trakl, *Das Fieberspital*)

> Streets that *follow like a tedious argument*
> *Of insidious intent*
> To lead you to an overwhelming question. . .
> (Eliot, *The Love Song of J. Alfred Prufrock*)

> Et moi j'ai le cœur *aussi gros*
> *Qu'un* cul de dame damascène. . . .
> (Appollinaire, *La chanson du malaimé*)

Similarly modern and surprising are the figures of speech in Donne (*Love's Growth*):

> And though each spring do *add to love new heat,*
> *As princes do in times of action get*
> *New taxes,* and remit them not in peace,
> No winter shall abate this spring's increase.

There is a qualitative difference between every-day, journalistic, and colloquial word associations and the linkage of words in poetry. A high degree of semantic entropy is a regular phenomenon in modern poetry, one of those essential features that distinguish poetic language from every-day language. Translators are aware of this and treat the syntagms and phrases of every-day language as a unit and substitute for them the most usual word groups in the target language. On the other hand, they translate the similes and unexpected turns of poetry word by word.

We can hardly help being aware that the language of every-day life is a different "language" from that of poetry when the poet deliberately reduces semantic entropy to the colloquial level. In these cases a predictable structure or phrase has the effect of a foreign word. It is with a scornful gesture

that Verlaine throws the clichés of polite language into the face of his divorced wife in the *Invectives* ("O, la Femme. . . "):

> Femme à jamais quittée, ô oui! *reçois* ici,
> Non sans *l'expression d'*un injuste regret,
> L'insulte d'un qu'un seul remord ramènerait.

He borrows the usual letter-closing formula *recevez l'expression de* from the *Parfait Secretaire* to change the stock phrase of apparent respect into a vivid expression of hatred. Ever since poetics has existed it has been known that a difference from the usual, a greater degree of entropy, is an organic part of poetry. Aristotle scorned Ariphrades who "used to ridicule the tragedians for introducing expressions unknown in the language of common life" because the beauty of poetic language is closely connected with the unusualness of the expressions. He shows with the help of eloquent examples how the beauty of a poem is gone as soon as we replace an unaccustomed turn with some redundant form.

In modern poetry, we find a variety of devices employed to exploit the relationship between accustomed and unaccustomed.

Word-pairs often used together because of the pleasing, often rhyming combination of sounds stand opposed to free word combinations. For modern poetry they often tend to acquire a startlingly new meaning through slight modification or appear in the shape of highly improbable combinations.

> *Maker and breaker*
> I am the ebb and the flood. . .
> (W. E. Henley, *I am the Reaper*)

> There will be time to murder and create. . .
> And time for a hundred indecisions,
> And for a hundred *visions and revisions*
> Before the taking of a toast and tea.
> (Eliot, *The Love Song of J. Alfred Prufrock*)

> The memory throws up *high and dry*
> A crowd of twisted things;
> (Eliot, *Rhapsody on a Windy Night*)

S de gyönyörű lenne	'What beautiful revenge it would be
Várva s zárva	*Waiting and closed*
Nevetni le a világra . . .	To laugh down at the world . . .

(Ady, *Csak rosszabb volnék* 'If I could be worse')

The dynamic effect of these pairs is due to the sharp contrast between the naïvely rhyming form and the extraordinarily free combination in meaning.

These pseudo-word-pairs are closely related to unusual alliterative word-combinations:

> That I may rise, and stand, o'erthrow me and bend
> Your force, to *break*, *blow*, *burn*, and make me new.
> (Donne, *Holy Sonnets*, XIV)

> alles wird um eine *Milde minder* . . .
> (Rilke, *Himmelfahrt Marias*, I)

> Demutsvoll beugt sich dem Schmerz der Geduldige
> Tönend von *Wohllaut* und *weichem Wahnsinn.*
> (Trakl, *In ein altes Stammbuch*)

> J'étais jeune je jouais comme un *rire respire*
> (Eluard, *L'enfance maîtresse*)

4. Small changes and cuts made on a first draft also reflect the poet's struggle against redundant word combinations. In the first version of the sonnet *Victorieusement fui* Mallarmé characterized the softly gleaming hair of a woman in a darkened room:

> Y verse *sa lueur diffuse* sans flambeau . . .

In the last version the line was changed to:

> Verse *son caressé nonchaloir* sans flambeau . . .

Mallarmé rejected the phrase *lueur diffuse*, which is suggested by *verse* and *sans flambeau*, cast off the non-essential adverb *y* and expended the six syllables thus gained in suggesting other aspects of the hair. The hair radiates "caressed indifference."

The first version of the *Pître châtie* began with the lines

> Pour ses yeux,—pour nager dans ces lacs, dont les quais
> Sont plantés de beaux cils qu'un matin bleu pénètre, . . .

The sonnet was published in 1887. By that time these two lines had been compressed into two words:[14]

> Yeux, lacs . . .

More mature works are usually characterized by less redundancy.

[14] *Œuvres complètes*, Paris, 1945, 31 and 1416. Of course, we do not mean to say that these two words contain for the reader all that Mallarmé explicitly wrote in the first version. For even a fair interpretation one must be familiar with Mallarmé's complete *œuvre*. Even so the question whether correct recoding would have been possible without having seen the first version remains.

The foregoing applies primarily, as the quotations indicate, to "modern" poetry from the second half of the nineteenth century on. But the war on conventional word-combinations began considerably earlier. Ronsard had already insisted on clipping banal and stereotyped attributes.[15] Striking adjectives that are peculiarly apt are also preferred by the English metaphysical poets.[16] As we pursue the development of medieval and modern European poetry, we must accept Charles Asselineau's statement that the law of gradual concentration held for poetry.[17] As the individual poet strives to reduce redundancy when polishing his poem, so the entropy of poetic communication increases in the course of the centuries.

5. The freedom of the speaker is restricted not only by transition probabilities, governed by grammatical and stylistic rules. The connection between words and objects determines what linguistic sign he is to use in a given situation. If we point to the hair of a woman as she combs it and ask someone what it is, he will answer in all probability: "Hair." Semantic rules that relate the signs to the things meant form the very basis of speech and social contact. Any reduction in transition probability makes understanding more difficult, while lifting the restrictions of semantic rules would make communication outright impossible.

It is interesting that despite this fact the poet struggles, and struggles successfully at that, against this kind of restriction. On the one hand words make communication possible, on the other hand they make the unspoiled interpretation of individual experience impossible.[18] The concept of hair includes all the common features of the hair of the "Recumbent Nude" by Modigliani, of a bank clerk with a receding hairline, of the adored and of the rejected mistress, and of an Australian kangaroo, i.e. what we can read under the entry *hair* in the *Oxford Dictionary*. The poet, on the other hand, has hardly ever anything to do with the *hair* in the *Oxford Dictionary*, with the general and universal concept of hair, the "Ding an sich," but he is interested in a "Ding für mich," concrete hair.

[15] Cf. H. Weber, *La création poétique au XVI^e siècle en France*. Paris, 1956, vol. I, p. 321.

[16] Cf. C. S. Lewis, *English Literature in the Sixteenth Century*. Oxford, 1954, pages 538ff.

[17] "A mesure que le livre imprimé, en se répandant, convertit les auditeurs impressionables, *passionables*, en lecteurs médidatifs et réfléchis, la poésie doit concentrer son essence et restreindre son developpement." Baudelaire, *Les Fleurs du Mal*. (Editions C. Lévy), Paris, 1922, Appendix, p. 320.

[18] Hegel has already pointed out that the basic contradiction in language lies in the fact that whereas it should express individual experience it can only express the universal. Cf. *Vorlesungen über die Geschichte der Philosophie*, 1833, XIV/2.

This unique primary experience cannot be conveyed or even suggested by a word *hair* which summarizes the common experience of millions of men and women, living and dead. Therefore the poet will invoke the assistance of the sentence to make the concept specific and to suggest the immediacy of experience.

Sometimes the poet asks for an even more radical solution. He ignores the word which has already existed before his own experience and is therefore, *subjectively* most inadequate. The word for him covers up and wipes out the experience. The rejected word is then replaced with an *objectively* inadequate word. Thus, *hair*, for instance, is replaced by *flame*, *cap*, *helmet*, *stream*, or *sea*. The relationship of these words to hair is unique and based on a single subjective experience. With the help of the context the reader understands that this word of another universal meaning is now used for hair. This method of indirect symbolization—called metaphor in rhetoric—conveys a large amount of information and does so not only because it is a word of poor predictability. The contradiction between the conventional basic meaning and the actual, situational meaning of the word *sea* or *flame* must be resolved. We must correct a statement that is, logically speaking, obviously incorrect. The "incorrect" term sets off an intellectual process which resolves the dissonance, sometimes very simply (by way of a simile) and sometimes in a more complex manner, with the help of a sequence of associations (often partly subliminal). The static connection between word and object yields to a dynamic, contradictory, dialectic form. (The sculptor depicting a galloping horse usually selects a link in the chain of movements which is the most remote from the position of equilibrium. In this way he makes it obvious that this movement must be directly followed by another movement lest the horse lose its balance. Thus the single movement caught by the artist already includes the next one.) The "false" sign denotes the object as if by detour and at the same time associates with it other objects approximated in the course of that detour (Figure 3).

> . . . dann giesst sie ihre Haare
> in den Spiegel . . .
> (Rilke, *Dame vor dem Spiegel*)

The verb "pours" evokes the image of a cascade of water pouring into a lake (the mirror) even before the concept of hair has been made clear.

Women's hair is a port which offers unparalleled security:

> Un port retentissant où mon âme peut boire
> A grand flots le parfum, le son et la couleur;
> (Baudelaire, *Le serpent qui danse*)

A flying flame:

La chevelure vol d'une flamme à l'extrême
Occident de désirs pour la tout déployer . . .
(Mallarmé, *La chevelure*)

A helmet giving out light in darkness from which the nude female body falls like a rain of roses:

Comme un casque guerrier d'impératrice enfant
Dont pour le figurer il tomberait des roses.
(Mallarmé, *Victorieusement fui*)

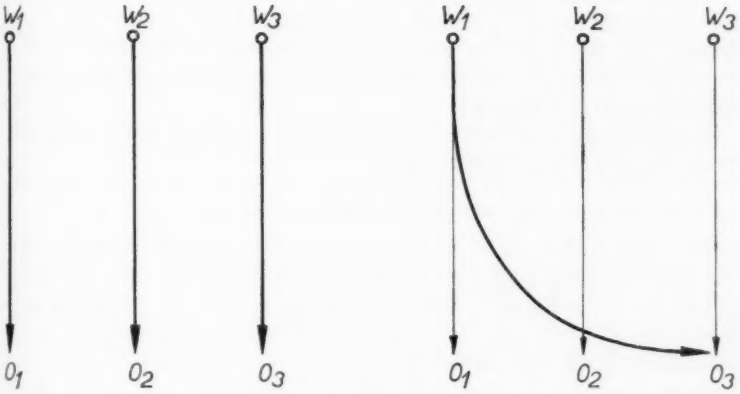


FIGURE 3. Model of denotation: *direct* (left) and *metaphorical* (right). The object is denoted indirectly by the "false" term (W_1) and identified with the right one (W_3) in context, without losing its original meaning.

7. Unique experience may be expressed by a newly created word, custom-made for that particular experience. Usually these newly-created words are universally understandable, for they are fashioned of commonly known elements according to the syntactic rules of the language. The inherent laws of some languages leave more room for new word formations than others. This is the case in German and Hungarian (as opposed to French and, to a lesser extent, English). There is hardly a poem by Rilke and none at all by Ady without such word-inventions.

Ich möchte dich wiegen und *kleinsingen*
und begleiten *schlafaus und schlafein.*
(Rilke, *Zum Einschlafen zu sagen*)

Szegény *hozzám-testvérisültek*, 'Poor *brothered-to me* [fellows]
Jámbor *élet-levetkezők* . . . Simple [souls who] *stripped life* . . .'
(Ady, *A halottak élén* 'At the head of the dead')

Kleinsingen 'to sing to one causing him to decrease in size'; *schlafaus und schlafein* is based on *tagaus tagein* 'day by day' and points to this accepted term, making it a part of the meaning of the new word, while it also applies literally, for it suggests accompanying the person falling asleep into her dream and then returning to the waking world. *Hozzám-testvérisültek* literally evokes the more common word combinations of *hozzámnőttek* 'grown close to me', *hozzámhasonultak* 'grown similar to me,' etc. for the Hungarian reader, and with these is associated the idea of becoming brothers. Thus the newly created word compresses the phrase "who have grown close and similar to me, who have become my brothers." *Élet-levetkezők* 'those who cast off life like a discarded garment,' refers to the dead of the First World War.

Sometimes the poet only gives a novel sense to an existing word, as Eliot does in a line already quoted:

And time for a hundred *indecisions* . . .

The numeral with the plural construction has an unusual effect because the word *indecision* originally does not denote single acts but a negative quality, "want of decision, want of settled purpose or of firmness in the determination of the will." Here again we are dealing with a transference of meaning, though the transference is more covert and between less remote concepts than in the case of metaphors. *Indecision* here means acts characterized by indecision. And yet it means more than just this, for it has brought with it something of its original meaning—among other things its abstract and impersonal quality and the very lack of action. All this corresponds very well to the meaning of the word as used in this context. This metaphor not only means but at the same time mimes indecision, vagueness, and passivity through the conversion of an attributive syntagm into an abstract substantive (with a negative mark).

The neologisms of Goethe or Heine are more transparent than the new words coined by modern poets:

Nach der Ebene dringt sein Lauf
Schlangenwandelnd. . .
(Goethe, *Mahomets Gesang*)

Die blauen *Frühlingsaugen*
Schaun aus dem Gras hervor;
(Heine, *Die blauen Frühlingsaugen* . . .)

Schlangenwandelnd 'serpentine-writhing; ' *Frühlingsaugen* 'blue eyes like spring flowers in the grass, like the sky in spring, eyes charming as spring time.'

Neologisms contain a larger amount of information than a single word and can be expressed, in everyday language, only in a sentence of a certain length. Poetic neologism shares uniqueness, the characteristic that it is applied only for a single occasion, with the sentence. This "word-sentence," however, appears in the real sentence organized into a unified concept, having given up its bipolar, predicative character. It denotes the concept, but at the same time says something new and very characteristic about it.

8. The poet wishes to convey his own thought—through his own words. The contradiction is inherent in this very aim. The concept or idea is no less universal and supra-individual than the word. Thinking built on concepts is also socially determined.[19] Entirely individual mental activity does not attain the level of real thoughts. By reducing semantic and grammatical redundancy to an unusual level, by relaxing linguistic and social restrictions, the poet attempts to communicate the incommunicable. This is the essence of poetic license. The thought expressed is accompanied by several other thoughts which are merely suggested. The increase in the amount of information goes hand in hand with the increase in noise. The implicit linguistic forms (semantic and grammatical metaphors, neologisms, sound colour, and rhythmic expression, etc.) are not interpreted with the same homogeneity by listeners and readers as are the traditional explicit symbols.

Experimental subjects, all university graduates, were told to define certain words in Endre Ady's poems. The participants, who easily defined the traditional words, said in the case of the newly invented words that they did not understand the word, only sensed that it had beauty and was expressive. When they were asked to explain the word informally as if by association, the distribution of the proposed meanings was wide but concentric.

Interpreters of certain lines written by Mallarmé often differ a great deal in their explanations. "It must be acknowledged," writes Guy Michaud on Mallarmé's poetry, "that, despite their sharp wit, commentators are still very far from being able to provide a satisfactory explanation for poems written in the 'latest style' (dernière manière)."[20] "In *Hommage à Wagner*

[19] Cf. L. Feuerbach, *Zur Kritik der hegelschen Philosophie*, [1839]; Berlin, 1955. pages 32f. —H. Wallon, *De l'acte a la pensée. Essai de psychologie comparée*. Paris, 1942, pages 91, 125, 195, *et passim*.

[20] *Mallarmé. L'homme et l'œuvre*. Paris, 1953.

he is almost incomprehensible, and it seems hardly possible to find agreement in the different explanations of different commentators" (*ibid.*, page 122). In case of a technical description it is most unlikely that the choice between a cup or a chandelier would remain vague ("Surgi de la croupe et du bond. . . "), whether we are reading about the division of reality and the soul, or about how to light a cigar ("Toute l'âme resumée").[21]

Some critics link Mallarmé's obscurity to his ambition to be concise:

". . . Il systématise davantage encore sa syntaxe, la rend plus hardie, plus concentrée, et, par une conséquence fatale, plus obscure; un bien plus grand nombre d'intentions s'accumule maintenant en une seule phrase."[22]

This is how the striving toward a greater amount of information, toward profound sincerity, and toward full communication free from distortions, lead to an obfuscation of the message. For a telecommunications engineer there is nothing astounding about this kind of contradiction, for he knows that the meaningless, incomprehensible white noise is the message with the maximal entropy, the greatest amount of information.[23]

9. This contradiction is rooted in the essence of communication. The contradiction we have found earlier is by no means irreconcilable. The great semantic entropy of poetic language stands in contrast to the predictability of its sounds. Of course, not even in the case of poetry can we determine the sounds of the word on the basis of its meaning. Nevertheless in the larger units of line and stanza, a certain relationship can be found between sounds and content. Dionysius of Halicarnassus (first century B.C.) already thought it important that there be harmony of form and content in a poem.[24]

[21] Cf. Michaud, *op. cit.*, page 168; Charles Chasse, *Les clefs de Mallarmé.* Paris, 1954, page 146. Or Aristotle: "The Diction becomes distinguished and non-prosaic by use of unfamiliar terms, *i.e.*, strange words, metaphors, lengthened forms, and everything that deviates from the ordinary modes of speech. But a whole statement in such terms will be either a riddle or a barbarism" or may achieve a comic effect (*De arte poetica liber*, chapter XXII, translated by I. Bywater).

[22] Jacques Schérer, *L'expression littéraire dans l'œuvre de Mallarmé.* Paris, 1947, p. 226.

[23] I. A. Richards considers the emotive use of words in poetry as the principal reason of ambiguity, obscurity of poetic messages. In the strict "symbolic language" of science the emotional effects of the words are irrelevant to their employment, whereas statements in poetry are there as a means for the manipulation and expression of feelings and attitudes. We have in symbolic (logical) language an appropriate apparatus for handling and elucidating sense. For handling feeling we have nothing at all comparable. *Practical Criticism* [1929], Harvest Books, New York, n.d., pages 177ff., 333. C. K. Ogden and I. A. Richards, *The Meaning of Meaning.* London, 1953, pages 149ff., 235ff.

[24] *De compositione verborum*, Lipsiae, 1808, Ed. Schaeffer, page 274.

The humanists Vida, Julius Caesar Scaliger, Opitz, Lomonosov, Diderot, Sully Prudhomme, and so many others all made it a requirement that the poet must not be satisfied with merely fitting his words to the thoughts he wishes to express, but he must also adapt the sounds of the poem to its content. But why should the poet use "harsh" sounds when speaking about rugged cliffs, and "small" ones when depicting small objects?[25] Why is this necessary if we can convey information much less ambiguously (at a lower noise level) with the help of conventional symbols, if it is possible to invoke images with mere sounds? How can this waste be reconciled with the principle of great economy observed in poetic communication?

A message conveyed with the help of words cannot be identical with a message communicated with mere sounds. Only conventional symbols can have meaning. The [ou] sound may suggest the sound of beasts, may invoke the images of darkness and mourning, /l/ and /m/ may have a soft effect, but none of them can substitute for the words *darkness*, *mourning*, *beast* or *softness*. The sounds are of interest to the poet probably just because of this imperfection. The essence of a symbol is that it points beyond itself, its own substance losing its significance, and is completely absorbed in its function. /č/ or /r/ have just as little to do with a chair as the graphs *ch* and *r* in the written form *chair* or the corresponding Morse-signs. The sounds we regard as "soft" or "dark" in a poem produce this effect under certain circumstances not through any convention, but through their material qualities. In the "nothing-but words" world of poetry the sounds salvage a fraction of the genuine experience, of reality. We get a plosive for an explosion, the /s/ sound for a storm, the /r/ sound for the stream of air from a brawl. The violation of metric laws is a result of Phèdre's irresistible passion. Now again the same tendency is seen at work as in the poetic application of words: to salvage in the poem as much original and individual experience as possible.

Of course, the principle of economy is at work, too. After all, the sounds play a double role. On the one hand they are elements in conventional symbols, connected with the content indirectly, through the word and the sentence; on the other hand they are immediately linked with experience through the natural relationship of sound and content. The difference between communication in everyday life and in poetry can perhaps be suggested by the diagram below (Figure 4).

If we compare this diagram with the model of the metaphor (Figure 3), a certain similarity of structure will be obvious. In both cases linguistic form approaches the object from two sides: That is how expression becomes

[25] Vida, *De arte poetica*, Rome, 1527, II, 403–412.

FIGURE 4. Model of "prosaic" and poetic denotation. Generally, sounds (S) are connected with objects (O) only through words (W) [left]. In poetic communication [right] sounds have a simultaneous direct link with experience.

plastic. Form is at the same time content. In poetic communication the message conveyed by sounds is in fact completed in the course of transmission. In the line from Attila József's poem *Esik* 'It is raining'

... dünnyög motyog a cipő ... '... the shoes mumble and mutter ...'

the "moist" palatals make the metaphor absolutely clear. The words have the meaning of *mumble* and *mutter*, as they are translated, but the Hungarian /ń/ and /ť/ sounds lost in the English translation give the original words a squelching and squashy quality, which really imitates the sucking sound of the wet shoes in the mud.

Sometimes the sounds give the lie to the words and differentiate the communication.

Arrière aussi les poings *cri*spés et la co*lè*re ...
(Verlaine, *Puisque l'aube grandit*)

Verlaine is speaking about reconciliation, but the line is full of hard /k/ and /r/ and hissing /s/ sounds. The polyphony of poetic communication makes possible the simultaneous expression of both poles of ambivalence.

The direct utilization of form must not weaken the primary mediatory function of form. As we have seen, investing the sounds with an independent function does not reduce the semantic entropy of poetry. Even though the /r/ sound has become more frequent than its average frequency, none of the words containing this sound seem to be superfluous or expendable. It seems to be impossible to express this idea with different words. (All this, of course, holds only for what we consider "really good poems," as

throughout this paper we have considered only works of very high quality.)

10. Thus, the redundancy of the sound elements of poetic language is only apparent. Nevertheless there are other redundant, verbose forms (in the every-day sense of the word) in poetic language: repetition of words, phrases, and lines, and various kinds of parallelism.

But there must be a qualitative difference between the redundancy of a poem and the redundancy of commonplaces; otherwise it would be incomprehensible that it is enjoyable in poetry and annoying in the case of clichés.

We would simply say that the redundancy of poetic language is in form and that of every-day language in content, but this is too vague an explanation to be satisfactory. The redundancy of poetic language is not really limited to linguistic forms in the strict sense of the term, or to sounds, but includes also the repetition of such elements of content as words and sentences.

Whether we consider rhyme, rhythm, or refrain, the striving for repetition is less restricted in poetry than in prose. This great redundancy is even more striking in comparison to the entropy that is equally characteristic of poetic communication, a sharp contradiction that does not exist in every-day language where the two are resolved in a compromise. Full repetition is prevented by the rules of style which are as strict as the rules of grammar. No linguistic or social convention obliges the speaker, however, to avoid predictable word associations. In contrast to the principle of *aurea mediocritas* operating in every-day speech, polarization, a centrifugal tendency, characterizes poetic language.[26]

In poetic language redundancy is directly and openly operative and is given special emphasis. This compulsion toward repetition (*Wiederholungszwang*) comes out in every-day communication by devious means. The speaker repeats not his own words but someone else's words, but without admitting this even to himself. He pretends to be expressing his individual opinion in his own words, formed after a careful weighing of the actual situation. Consider the statesman who reads a "spontaneous" message off a slip of paper at the airport, the perennial children's answer "nothing special" in answer to the adults' inane "What happened at school today?"

[26] I. Hermann found in experiments with small children that, in contrast to adults, they are more likely to choose the object on either end of a row rather than the object in the middle. He concludes that this centrifugal tendency is an integral part of the primitive intellectual mechanism (*Primärvorgang*). "Die Randbevorzugung als Primärvorgang," *Internationale Zeitschrift zur Psychoanalyse*, IX, 1923, pages 137ff.

and our contentless "How do you do?" which does not require an answer, or even our antipathy toward pre-fabricated "witty sayings," stale jokes, and puns which become more irritating the greater the gap between pretended and actual wit.

There is another way in which we are wronged by banal stereotypes. Open playful repetition may be freely enjoyed by the reader of poetry. Our highly critical inner judge, the superego which pronounces stricter and stricter sentence on the pleasure of monotonous repetition beginning with the age of five or six, is more indulgent in this case, for the repetitions are obviously an intellectual product of a high order. The poet apparently pays for the primitive, childish joy of repetition in concrete, highly differentiated and highly informative linguistic form which is of social value. In every-day colloquial speech the constant change of form covers up for the repetition of the words of others. In poetry the very opposite happens. What seems to be repetition often hides new meaning. An unchanging refrain gains a different shade of meaning, and sometimes an entirely different meaning, in every successive stanza. Let us take Villon's classic refrains "*Mais où sont les neiges d'antan*," "*Bien est eureux qui riens n'y a!*" "*Il n'est bon bec que de Paris*," or even the refrains of *chansons* whose meaning may change in the last stanza often to the very opposite of the original.

In the first stanza of Verlaine's *Chevaux de bois* the word *tournez* is found eight times.

> *Tournez*, *tournez*, bons chevaux de bois,
> *Tournez* cent tours, *tournez* mille tours,
> *Tournez* souvent et *tournez* toujours,
> *Tournez*, *tournez* au son des hautbois.

The word occurs twice in every line, in two lines in unbroken repetition, and twice introducing the two halves of the line. Thus even the very mode of repetition is restricted, following the formula *abba*. At the same time, however, this repetition has a message of its own. It depicts the movement of the merry-go-round, it brings to mind the wooden horse as it makes its rounds. The rhythm of the poem is also extensively functionalized and put into the service of communication. The restriction of the hexameter did not prevent Vergil from slowing down the rhythm to portray slow movement and suggest effort,[27] as was already recognized in classical rhetoric.[28]

Less has been said about rhymes contributing to the content.

[27] P. Vergilius Maro, Æneis Buch VI, erklärt von E. Norden, Berlin, 1916, 4323.

[28] Dionysius of Halicarnassus, *De compositione verborum*, pages 275ff. J. C. Scaliger, *Poetices libri septem*,3 1586, p. 489. Diderot, *Salons de 1767*, *Œuvres complètes* XI. Paris, 1876, p. 331.

In *Faust*, Part II, Act II, the threefold double rhyme varied with inner rhymes suggests the honeyed tones of Phorkyas-Mephisto.

Buchstabirt in Liebes-Fibeln,
Tändelnd grübelt[29] nur am Liebeln,
Müssig liebelt fort im Grübeln[29] . . .

The monotonously repeated rhymes may suggest the fearful quiet of the evening:

Wo du gehst wird Herbst und Abend,
Blaues Wild, das unter Bäumen *tönt*,
Einsamer Weiher am Abend.

Leise der Flug der Vögel *tönt*,
Die Schwermut über deinen Augenbogen.
Dein schmales Lächeln *tönt*.
(Trakl, *Rosenkranzlieder*)

Generally, even in classic French poetry, with its modest means, the rhymes stress internal connections. It is not by chance that Corneille likes to rhyme *bonheur* with *honneur*. Heine's scornful grin is evident behind a rhyme in *Tambourmajor*, where he connects the nationalist Körner with horns by means of the rhyme:

Wie in der Kampfbahn der Auerochs
Erhuben wir unsere *Hörner*,
Entledigten uns des fränkischen Jochs
und sangen die Lieder von *Körner*.

In Roman Jakobson's opinion "words similar in sound are drawn together in meaning." "Any conspicuous similarity in sound is evaluated in respect to similarity and/or dissimilarity in meaning."[30] Even the playful rhyming of lines can therefore be called into the service of communication. The playful form of a poem is the result of complex (and mostly preconscious) effort. What intellectual effort is spared the reader is proportionate to the measure of the intellectual achievement of the poet. The more informative, more economical, and more enjoyable this form is for us readers, the greater the intellectual feat behind it. On the other hand, the author of commonplaces saves intellectual energy at our expense. It is understandable then that society evaluates the two kinds of redundancy by a double standard.

11. All this is in full accord with the basic attitude of the poet: he, rather than others, enunciates things that are difficult to say, and thus frees

[29] The *ü* in *Grübeln, grübelt* was pronounced as a short *i*.

[30] 'Linguistics and Poetics' in *Style in Language*, edited by Thomas A. Sebeok, New York and London, 1960, pages 371, 372.

society—and himself—from tension caused by conflict. With his economy in the form of poetry he saves the public intellectual effort and with the content of poetry he reduces the role of inhibition (*esparter Hemmungsaufwand*).[31]

Poetry can achieve this liberating effect only if it husbands its word-material with maximum care. Only concrete,[32] accurate and completely personal expressions can count on general recognition and success. But if the freeness and entropy of linguistic form stands in the service of expression, then the contradictory tendencies, the striving for maximal entropy and great redundancy, are resolved. To *express* means, in the original sense of the word, to push something out. We must indeed expel the stimulant which causes the tension in order to regain a normal state of equilibrium. Thus, the poet combats the every-day manifestations of lax, redundant forms—the every-day manifestations of the compulsion for repetition—so that he may let his own compulsion triumph on a higher esthetic level."[33]

The repetition of syllables, words and lines, the cuts in the "cost" of communication and of intellectual expenditure for increasing entropy and better expression is the manifestation of the same basic tendency on different levels. But the road is very long from monotonous repetition to the most effective forms of expression: the millennia-old road of domesticating and socializing primitive instincts. Their common origin does not reduce the fundamental difference between the two.

Department of Phonetics
Institute of Linguistics of the Hungarian Academy of Sciences
Budapest V
Szalay utca 10

[31] Cf. O. Rank, *Der Künstler und andere Beiträge zur Psychoanalyse des dichterischen Schaffens*. Leipzig–Wien–Zürich, 1925, pp. 52f; S. Freud, *Der Witz und seine Beziehungen zum Unbewussten*. (*Gesammelte Werke*, vol. VI), London, 1940, p. 133.

[32] The word *concrete* is here used in its philosophical rather than in its every-day sense. A given object is necessarily concrete, but not only tangible objects are concrete. Everything is concrete that is the occasional focus of different properties (*Einheit unterschiedener Bestimmungen*). Cf. Hegel, *Encyklopædie der philosophischen Wissenschaften*, 82.

[33] S. Freud, *Jenseits des Lustprinzips*. (*Gesammelte Werke*, vol. XIII), London, 1940, p. 67f.

Discussion

SOME METHODOLOGICAL REMARKS ON GENERATIVE GRAMMAR

NOAM CHOMSKY

1. These remarks are prompted by several critical discussions[1] of recent work in descriptive syntax and syntactic theory carried out within the framework of generative grammar.[2] I will try to clarify several issues that have, apparently, been the source of some misunderstanding and confusion as to the purposes and assumptions of this work, and to indicate how it can perhaps be extended into certain important but neglected areas.[3]

It is useful, at the outset, to make a distinction between *data* and *facts*. The linguist's data consist of certain observations about the form and use of utterances. The facts of linguistic structure that he hopes to discover go well beyond these observations. A grammar of a particular language is, in effect, an hypothesis about the principles of sentence formation in this language. It represents a factual claim concerning the rules that underlie the data that have been collected. We judge the truth or falsity of this hypothesis by considering how well the grammar succeeds in organizing the data, how satisfying an explanation it provides for a wealth of empirical observations, how far-reaching are its generalizations, how successfully it accomodates new data. An enormous amount of data is available to any native speaker; the deeper facts of linguistic structure, however, remain hidden to him.

Similarly, a theory of linguistic structure is an hypothesis about linguistic universals. It asserts that the grammars of all human languages are constructed on such-and-such a specified plan. Such a theory should explicitly

[1] Hill, "Grammaticality," *Word* XVII (1961) (referred to henceforth as [G]); Jakobson, "Boas' view of grammatical meaning," *The Anthropology of Franz Boas, American Anthropologist* (1959) (henceforth, [BVGM]); Bolinger, "Linguistic Science and Linguistic Engineering," *Word* XVI (1960) (henceforth, [LSLE]).

[2] As exemplified, in particular, in my *Syntactic Structures*, The Hague, 1957 (henceforth [SS]).

[3] Much of what I say here is a paraphrase of earlier remarks, e.g., in [SS] and in "A transformational approach to syntax," *Third Texas Conference on Problems in the Analysis of English*, A. A. Hill, editor, forthcoming (henceforth, [TAS]).

characterize the form of grammars (the schema for grammatical description) and should provide a method for selecting among alternative grammars, each compatible with the available data from some language (and each making a specific set of factual claims concerning this language). A theory of this sort can be verified in an indirect, but quite satisfactory way. We can ask whether, for each language, the most highly valued grammar of the specified form is confirmed in the sense described in the preceding paragraph.

A linguist who is exclusively concerned with linguistic data and not with the facts of language, in this broader sense, has severely limited the scope of his research.[4] A grammatical description that gives nothing more than "a compact one-one representation of the stock of utterances in the corpus"[5] can be challenged only to the extent that the observations it summarizes or rearranges are defective. It claims little, and its interest is correspondingly limited.

In [SS] I was concerned with grammars that make a much stronger factual claim than this. A *generative grammar*, in the sense of [SS], is not a large collection of neatly organized examples, supplemented with comments about these examples and hints as to how to construct similar ones. Nor is it a discussion of efficient and compact notations (e.g., inventories of phonemes, morphemes, categories or construction types) in terms of which the utterances of a corpus can be represented. A generative grammar is a system of explicit rules that assign to each sequence of phones, whether of the observed corpus or not, a structural description that contains all information about how this sequence of phones is represented on each of

[4] In his [LSLE], Bolinger distinguishes between data-oriented and model-oriented approaches to linguistic description, the implication being that 'model-oriented' studies are less concerned with the facts of language as such. The essential difference between the two approaches that Bolinger describes seems to me, rather, that his 'data-oriented approach' limits its concern to a narrow subset of more immediately accessible facts; it does not attempt the precise formulation of deeper underlying generalizations.

It should be remembered, in this connection, that the most partial and limited set of explicitly formulated recursive (iterative) rules is, literally, far more extensive in coverage than the most immense collection of examples assembled and arranged by the data-oriented linguist. We return to this matter below.

[5] At one point (p. 366) in his *Methods in Structural Linguistics* (Chicago, 1951), Harris describes this as "the over-all purpose of work in descriptive linguistics." Later, however, he remarks (p. 372) that "the work of analysis leads right up to the statements which enable anyone to synthesize or predict utterances in the language." It is this set of statements that I am calling the grammar. I am not convinced that the last-quoted remark is really justified (that is, that any known procedures do provide an adequate grammar), but this is another question. At the moment I am only concerned to make the difference in aim explicit.

the several linguistic levels—in particular, information as to whether this sequence of phones is a properly formed or *grammatical* sentence[6] and if not, in what respects it deviates from well-formedness.[7] In particular, then, this grammar distinguishes a class of perfectly well-formed (fully grammatical) sentences. It is designed, in other words, to meet what Hockett has proposed as the basic test of significance for a grammar, namely, that it "generate any number of utterances in the language, above and beyond those observed in advance by the analyst—new utterances most, if not all of which will pass the test of casual acceptance by a native speaker."[8] To Hockett's remark I should only like to add that a grammar must not merely specify an infinite class of properly formed utterances, but must also assign to each sequence of phones a structural description that provides a basis for explaining how this utterance is used and understood—that provides the structural information without which it is impossible to undertake this further task in a serious way.

Three theories of generative grammar are proposed for comparison and study in [SS]. The first (based on the notion of finite state Markov process or finite automaton) was shown to be too weak. The second was a formalized and somewhat generalized version of the theory of immediate constituents, reinterpreted as a theory of generative grammar. The third was a transformational model. I tried to show, by considering several kinds of linguistic data, that the immediate constituent theory was inadequate, and that the transformational theory seemed much more promising. Additional

[6] In grammars of the type studied in [SS], part of the structural description of a phone sequence is its representation on the phrase structure level by a set of abstract strings that can be presented, equivalently, as a tree with labelled nodes. See [SS], and references there, for details. The fully grammatical sentences generated by the grammar are the ones represented by trees headed by a single node labelled *S*, where *S* is the distinguished initial symbol of the underlying constituent structure grammar. Transformational rules also play a part in assigning a phrase structure representation, in that theory, although the question exactly how they do so has not yet been satisfactorily answered. For discussion of this see my *Logical Structure of Linguistic Theory* (mimeographed, 1955, henceforth, [LSLT]), of which [SS] is essentially an excerpt, and "On the notion 'rule of grammar'," *Proceedings of the Symposia on Applied Mathematics*, XIII (1961).

[7] That is, sentences not generated directly can still be assigned a degree of grammaticalness, and a structural description which will indicate how they are related to directly generated sentences. A chapter of [LSLT] is devoted to this question, which is mentioned only in passing in [SS] (footnotes 2 and 7 of chapter 5). We return to this below in §5.

[8] "Two Models of Grammatical Description," *Word* X (1954), reprinted in Joos, ed., *Readings in Linguistics* (Washington, 1957). See also Hockett's "A Note on 'Structure'," *International Journal of American Linguistics* XIV (1948), also reprinted in Joos' *Readings*.

work that has appeared since then[9] seems to me to give support to this claim, which, I should like to emphasize again, is a factual claim concerning English and, more generally, the structure of natural languages—a claim that can be proven false in several ways; e.g., by the construction of grammars of a different form that are more successful in handling data of the kind that originally motivated the theory, by a proof that any grammar of this form must fail, for some natural language, because of certain formal features of utterances in this language, etc.

2. The investigations of generative grammar described in [SS] were motivated in part by an interest in the problem of accounting for the ability of a speaker to produce and understand an indefinite number of new sentences (or for that matter, to recognize them as properly formed, or as deviating from well-formedness in one or another respect) ([SS], p. 15), a task that he performs regularly with great facility. A generative grammar can be regarded as an attempt to characterize certain aspects of this ability, and a particular theory of generative grammar is a proposal concerning its general and universal features.[10] If we hope to make significant progress towards these goals, the grammars we construct must be sufficiently precise so that we can deduce from them consequences that can be confronted directly with linguistic data of a rich and varied sort. This requirement is rarely met by either traditional or modern studies,[11] though this fact may be obscured by the informality of linguistic descriptions, whether traditional or modern.

The study of precise generative grammars with explicit consequences

[9] E.g., [TAS] (which is, again, largely an excerpt from [LSLT]) and Lees' *Grammar of English Nominalizations* (Baltimore, 1960), henceforth referred to as [GEN].

[10] A theory of generative grammars can be regarded quite naturally as a proposal concerning certain fundamental and specific skills that the child brings to language learning. See my "Explanatory Models in Linguistics," *Proceedings of the International Congress on Logic, Methodology and Philosophy of Science* (August 1960), to appear, P. Suppes, editor (henceforth [EML]), for some further discussion.

[11] With some very limited exceptions, e.g., inflectional paradigms of the traditional sort. Given a list of lexical items, a paradigm gives explicit instructions for constructing a finite set of inflected forms. But it is clear that this is a relatively minor success.

It is on the level of syntax, of course, that the intuitive character of grammatical descriptions, and the defects of such descriptions, are most apparent, since it is here that the data cannot be enumerated in a finite list. However, I think that the preoccupation of modern linguistics with inventories rather than rules has had the effect of removing even significant aspects of phonology from the domain of linguistic research; for example, questions of the kind mentioned further below. For further observations on this see [EML], and Halle and Chomsky, *Sound Pattern of English* (forthcoming, henceforth, [SPE]).

(and the general form of such grammars) is in its infancy. Traditional grammars rely on the intelligence and linguistic intuition of the reader to provide the structural descriptions of items that do not appear in the collection of examples. There is little doubt that deep features of grammatical structure are unexpressed in a grammar addressed to an intelligent human who, in some totally unknown way, acquires the ability to produce and understand new sentences, having absorbed the information presented to him in a compendious traditional grammar. The study of generative grammars is, however, a natural outgrowth of traditional descriptive linguistics. Modern linguistics has, typically, been concerned with the much narrower problem of constructing several inventories of elements in terms of which utterances can be represented, and has given little attention to the rules that generate utterances with structural descriptions.[12]

3. Clearly if our goals are as described in §2, we will bring to bear data of many kinds in attempting to determine the validity of particular generative grammars or, more seriously, particular theories of generative grammar. Thus in [SS], [TAS], and elsewhere, I have been concerned with such data as

(1) (a) phonetic transcriptions;
(b) judgements of conformity of utterance tokens;
(c) judgments of well-formedness;
(d) ambiguity that can be traced to structural origins;
(e) judgments of sameness or difference of sentence type;
(f) judgments concerning the propriety of particular classifications or segmentations;

and so on.[13] By (a), I refer to the kind of data given by impressionistic

[12] In [TAS] I noted a few exceptions to this. In particular, Harris' morpheme-to-utterance rules can be reinterpreted as a model for generative grammar, and Hockett's Markov process model was explicitly designed as a sentence-generating grammar.

[13] To avoid a possible misunderstanding, notice that I am suggesting that data of these kinds are relevant to determining the validity of grammars and linguistic theory, but not that these data should be utilized in the construction of or choice among grammars, insofar as this is specified by the theory of linguistic structure. The difference is fundamental. See [SS] and [EML] for further discussion. Thus we might try to develop a general theory of linguistic structure powerful enough to lead us from certain data about a language to a generative grammar of this language, limiting the data, in this context, to (a), (b), (c), that is, to the kind of data that we might regard as being available to the child who learns language. This might be a reasonable decision for someone whose motivation in studying generative grammars is to construct some sort of model for language learning. But such a person would still concern himself with such data as (d), (e), (f) in evaluating the success with which his theory of language learning duplicates the performance of the human.

phonetics of the usual kind, e.g., transcriptions of stress contours, etc. Category (b) includes the intuitive sameness-difference judgments that are fundamental to descriptive phonology, e.g., the observation that in my normal speech, *latter* and *ladder* are phonemically identical (i.e., tokens of them are) but *writer* and *rider* are phonemically distinct. To illustrate (c), I gave such examples as

(2) colorless green ideas sleep furiously
(3) furiously sleep ideas green colorless
(4) the child seems sleeping
(5) the book seems interesting

and several others. Thus I am as certain of the fact that (2) and (5) are well-formed (in a sense in which (3) and (4) are not) as I am of particular impressionistic phonetic transcriptions (e.g., of the various *light house keeper's*) or of particular judgments of sameness and difference. In connection with (d), I was concerned with such forms as *flying planes can be dangerous*, *I found the boy studying in the library*, etc. Examples of type (e) are of many kinds. Thus I discussed the problem of finding a grammatical explanation for the fact that *did you eat the apple*, *who ate the apple*, *what did you eat*, etc. are all naturally regarded as belonging to a special sentence type (as distinct from *you ate the apple*, etc.), and the problem of explaining the intuitively obvious difference between *the police suspected the man behind the bar*, . . . *questioned the man behind closed doors*, . . . *put the man behind bars*. Many other questions of the same kind were also raised, e.g., why *smôking mén* seems to mean "men who are on fire" rather than "men who smoke." Concerning the propriety of segmentation and classification, it would be absurd not to use the fact that

(6) [it ended late ([iDêndid+léyt], [D]=alveolar flap)

obviously contains at least four morphemes in evaluating a theory of morphemic analysis, or not to use the fact that *boy*, *horse*, *justice* are, in a certain clear sense, words of a different type from *compel*, *retain*, *bring*, etc., in determining the success of a theory of word classes.

Considerations of all of these kinds are familiar in traditional grammar. Considerations of kinds (a), (b), (f) are explicit in modern structural linguistics. Thus the methodological studies of American descriptivists are motivated by information of type (f), presumed known. If someone were to propose a theory of morphemic segmentation that led to the conclusion that (6) contains two morphemes, this consequence would, presumably, be regarded as grounds for rejecting the theory. The same is true of proposed methods (e.g., substitution procedures) for finding classes, for defining the phoneme, etc. These methodological studies (and the controversies to

which they give rise) become utterly unintelligible if we do not make the perfectly justified assumption that they are an attempt to characterize precisely something that is, in part, intuitively given. The attempts to define 'phoneme', 'morpheme', etc. presuppose a set of clear cases of applicability or non-applicability of these terms, with reference to which the adequacy of the proposed definition can be tested. The same is true of an attempt to define 'well-formed (grammatical) English sentence' (that is, the attempt to construct an English grammar), or the attempt to define 'well-formed (grammatical) sentence in an arbitrary language L'. In terms of the preceding discussion, these are attempts to discover and state the facts that underlie and account for the observed data. A refusal to use data such as (1) would eliminate linguistics as a discipline, just as surely as a refusal to consider what a subject senses or perceives would destroy psychophysics. In both cases, we are trying (though in very different ways) to find a basis for intuitive judgements. In both cases, furthermore, the difficulty of obtaining reliable and relevant reports is quite apparent.

It seems to me perfectly plain that an enormous amount of data of this sort is available to the native speaker (though it does not follow that it is easy to elicit such data). It is pointless to refuse to make use of such data in evaluating the success of a particular attempt to formulate the rules of some grammar (i.e., the rules that the native speaker must somehow have internalized when he has achieved the ability to produce and understand new sentences) or the principles that underlie grammatical structure in general.

It is data of the kinds mentioned above, it seems to me, that we are trying to characterize precisely, to account for and reduce to general underlying principles, when we study grammatical structure. A generative grammar can be thought of as a theory of the language that attempts to characterize and explain data of this sort. This point of view is discussed at length in [SS] (e.g., §§6, 8), [EML] and elsewhere, but I will illustrate what I have in mind by a few examples.

In the case of data of type (1a), a generative grammar will attempt to exhibit general rules that underlie particular phonetic observations. It will try to show how the stress contours of English phrases and words are related to and determined by their immediate constituent structure and segmental phonemes;[14] it will offer an account of vowel reduction (as in

[14] Cf. Chomsky, Halle, Lukoff, "On Accent and Juncture in English," *For Roman Jakobson* (The Hague, 1956). A new version of this material, simplified, generalized, and incorporated into a much more general study of phonological processes, will appear in [SPE]. It is described in part in [EML] and (in an earlier version), in my contribution to the IV[th] Texas conference (1959).

the second syllable of *compensation*, but not *condensation*, in my speech) in terms of IC and syllable structure (cf. [EML] and [SPE]); it will try to explain why such items as *telegraph*, *aristocrat* (in the contexts #—#, —+y, —+ic), etc., have a particular set of variants, instead of others, by showing that this is a result of application of independently motivated rules (cf. [EML] and [SPE]); and so on, in innumerable similar cases. A generative grammar will be judged successful—that is, it will be judged to have brought to light the underlying facts of linguistic structure—to the extent that it shows how a relatively small number of simple and systematic rules account for a substantial quantity of such data. Cf. §1, above.

Similarly, in the case of data of the kind (1c), the generative grammar will attempt to exhibit interconnections among apparently independent judgments as to well-formedness and to state general rules from which a large variety of such judgments (e.g., those mentioned above (6)) can be deduced as consequences. Cf. [SS], [TAS], [GEN], for many examples.

There is a second and independent way of approaching the study of data of kind (1). We can try to discover objective experimental techniques for characterizing these intuitive observations. Thus in the case of impressionistic phonetic observations of type (1a), we can try to discover objective acoustic or articulatory correlates to what the linguist hears as stress level, vowel quality, and so on. And in the case of data of kind (1c), we can try to construct behavioral tests that will correspond, in their results, to the intuitive judgments of well-formedness.[15]

Here, then, are two approaches to the data of linguistics. We can attempt to characterize the data by a generative grammar or by experimental tests. The two approaches are not alternatives. Obviously, both are justified and we should like to pursue them in such a way that they give convergent results. It is important to realize that each approach complements the other and that neither presupposes the other. Whichever approach we are pursuing, the criterion of success is provided by the data

[15] Linguists have in general not been much concerned with this approach to linguistic data. Thus even in the case of (1b), where a reasonably good test has been suggested (namely, the paired-utterance test proposed by Harris, op. cit., pp. 32f.), there is not a single paper in the linguistic literature devoted to an objective evaluation of its results, despite the fact that the relation of conformity (sameness-difference) among utterance tokens has been regarded, since the publication of Bloomfield's *Postulates*, as the cornerstone of structural linguistics. There is no doubt that this test will often fail to give the right results, in the form in which it is usually described. Some ways in which it can be improved have been suggested (e.g., in [SS], p. 96), but have not been systematically investigated.

that give rise to the investigation; in this case, such data as (1).[16] Suppose, for example, that someone proposes that impressionistic judgments of stress level correspond to physical measures of intensity. He will quickly discover that 'stress' as so defined does not at all correspond to the impressionistic transcriptions that constitute the phonetic record (1a). In the face of just this evidence, it would be absurd for him to conclude that the transcriptions are incorrect, that phonology is built on sand, etc. This lack of correspondence may simply show that he has chosen the wrong test.[17] Similarly, if the paired utterance test, administered in a certain way, fails to show that *writer* and *rider* are phonemically distinct in my speech, this is evidence that the proposed test requires revision. Or, if a test for grammaticalness (i.e., a test of (1c)) is proposed which fails to make the required distinction between (2), (5) on the one hand and (3), (4) on the other, we can conclude only that the test is poorly constructed.

In other words, we evaluate the success and relevance of an operational test (just as we evaluate the success of a generative grammar) by asking how well it corresponds to the given data. There is no difficulty in constructing a bad grammar. It is just as easy to construct a bad test. There is no interest in either.

4. I would like to turn now to some of the criticisms of [SS]. Consider first Hill's [G]. What Hill shows is that several tests that he has tried are quite bad. I do not see the importance of these negative results, nor do I see any

[16] To some extent, of course, theoretical investigations or operational tests may lead us to reject some of the original data as irrelevant, or as probably incorrect or contaminated. We can expect that one result of systematic study, of either kind, will be to improve our understanding of what are the natural bounds of the discipline or what phenomena are relevant to deepening our investigation of linguistic structure.

[17] Continued failure of attempts to find a test might, ultimately, shake our confidence in the phonetic record, but it is not easy to say just when this would be justified. Thus we might discover that what the phonetician hears as stress level is determined in part not by the acoustic event but by an ideal stress pattern that is assigned to the utterance by the rules of a generative grammar—that is, the phonetician may be "hearing" something that is not present in the sound wave in full detail, but that is implied by the grammatical structure of the sound wave that he has understood in a certain way. I think, in fact, that this is not at all out of the question, in the case of stress levels. Cf. [SPE].

Similarly, a continued failure to elicit the judgment that (2), (5) are distinct in some significant way from (3), (4) might ultimately shake our confidence in this particular judgment, at least for certain speakers. It is, I suppose, conceivable that some speakers of English are unaware of the difference between (2), (5) and similar examples, on the one hand, and (3), (4), (or "tall the man cigar the smoked black," with whatever intonation—cf. [G], footnote 4), etc., on the other. Such a person, if he exists, is likely to be about as successful in studying syntax as a deaf man would be as a phonetician.

reason to find them at all surprising.[18] Thus if we ask a speaker the question "which of these sentences are grammatical?" giving him no idea how this question is to be interpreted, we will naturally expect the responses to be of no use in characterizing data of type (1c). Similarly, if we were to ask a speaker "what do you hear?" his responses would probably be of little phonetic relevance (they would have little relation to (1a)); or if we ask "were these utterances the same or different?" not indicating how we intend this question, we would expect the result to correlate poorly with the classification of utterance tokens as phonemically identical or distinct (cf. (1b)). Lack of correspondence, in such cases, demonstrates nothing about the validity of such observations as (1a), (1b), (1c). It merely indicates, as we should have expected in advance, that more subtle means will be necessary if we are to obtain relevant information about the speaker's awareness.[19]

Hill does suggest one test which might yield some useful data. In this

[18] Hill seems to have the impression that I have suggested these tests as general criteria of grammaticalness or degree of grammaticalness. But this is not the case. In [SS], I made it clear at the outset (p. 13) that I was not proposing any general behavioral test. I did assert (p. 16, as an aside, in the context of a discussion of order of approximation) that a speaker of English will treat (3), but not (2), as a sequence of unrelated words, assigning phrase final intonation to each, and that he will be able to recall (2) more easily than (3), to learn it more quickly, etc., despite the fact that they are equally new to him. I was careful *not* to suggest that these remarks offer a general criterion for grammaticalness—they would, for example, no doubt fail in the case of (4). (5). Later, on pp. 35–36, I made some further remarks on phonetic accompaniments to deviation from grammaticalness, again suggesting no general operational criterion. In fact, I know of none.

There is only one apparent conflict between [SS] and Hill's results, namely, he has found that his informants did not assign phrase final intonation to each word in (3), as I did. This lack of correlation is not at all surprising. In such quasi-operational tests as these, the instructions, the setting and so on can have a significant effect on what the subject does. It is for this reason, among others, that neither my comments nor his can be taken seriously as the description of an experimental procedure. Talk of "statistical reliability" ([G], p. 7) in connection with such informal suggestions as these betrays a misconception as to what is a serious scientific experiment. In any event, the question how the particular sequence of words (3) is read (nothing else is at stake) is surely too insignificant to justify a sustained discussion.

[19] This is noted explicitly by Maclay and Sleator in their "Responses to Language: Judgements of Grammaticalness," *International Journal of American Linguistics* XXX (1960), 281. Their tests apparently gave somewhat better and more relevant results than the one that Hill describes. In his "Language, Purposive Behavior and Evolution" (mimeographed, 1958), Eric Lenneberg describes a test of grammaticalness applied to (2) and randomly chosen rearrangements of the words of (2) which, given to M.I.T. undergraduate language students with appropriate instructions (there stated), gave the expected results unequivocally.

test, two sets of utterances are presented to a subject who is asked, essentially, whether the same formal relation holds in the two sets. Hill regards this test as "more structural" than a direct test of grammaticalness, and considers it somehow more relevant to the "transformationist."[20] Insofar as I can reconstruct the reasoning that leads to this conclusion, it seems to me to be based on two confusions. The first has to do with the nature of transformational relations. Thus some of the sets that Hill presented in his test contain sentences that are transformationally related to one another, but others do not. If Hill considers this test particularly relevant to transformational grammar, it may be in part because he has misinterpreted the technical use of the term "grammatical transformation" in [SS], [TAS], [GEN], and elsewhere, taking it to refer to any formally statable relation among sentences. I do not think it has ever been used in this sense.

Secondly, there is apparent in Hill's paper a confusion about the relation between the terms "generative grammar" and "transformational grammar." Hill seems to identify the two. Similarly, Bolinger regards Jakobson's comments about grammaticalness (to which I return below) as having some particular relevance to "transformationalists" ([LSLE], p. 377). In fact, neither Hill's nor Jakobson's comments have any direct bearing on the transformational model for generative grammar. They apply rather to generative grammar as such, that is, to the attempt to construct precise grammars with explicit consequences as to well-formedness of utterances. Grammaticalness is no specific concern of those who happen to think that the particular transformational theory proposed in [SS], [GEN] and elsewhere is the most promising specific characterization of generative grammars. It is, rather, the concern of anyone who is interested in syntax in the traditional sense, or in the problems discussed above in §2.

One final set of criticisms of transformational grammar presented in [G] also requires some clarification. Hill is under the impression that those of us who are exploring transformational models for generative grammar believe that sentences are primarily to be regarded as strings of printed words, and that we reject the view that "language is basically a system of sound, only secondarily one of writing." This, he believes may "explain the hostility of many structuralists." In [SS], [GEN], and elsewhere, a

[20] It should be observed that this test, as presented, is not yet very useful. Thus it does not tell us which of the two sentences "the child is sleeping," "the child seems sleeping" is to be rejected as deviating from well-formedness. Furthermore, it does not seem to be generalizable to such distinctions as (2)–(3). However, it can perhaps be refined, and it should be considered as one of a battery of tests that may ultimately succeed in giving a characterization of grammaticalness.

grammar is regarded as consisting of a syntactic component and a morphophonemic component. The former generates an infinite class of strings of minimal syntactically functioning units. Each such string, furthermore, automatically receives a structural description. The morphophonemic component converts each item given as output by the syntactic component into, ultimately, the phonetic representation of an utterance. Both [SS] and [GEN] are devoted primarily to the syntactic component. Other studies have been primarily devoted to the morphophonemic component of a generative transformational grammar.[21] The fact that the output of a generative grammar is a representation of speech is made quite clear in [SS] (e.g., p. 46) and [GEN], however.[22]

Jakobson, in his [BVGM], also indicates his dissatisfaction with the use to which such examples as

(6) colorless green ideas sleep furiously
(7) furiously sleep ideas green colorless

were put in [SS]. This criticism (which is cited with approval in Bolinger's [LSLE]) I find quite curious and puzzling. What Jakobson says about these examples, and about the question of grammaticalness and degree of grammaticalness in general, seems to me in full support of the position maintained in [SS]—in fact, it is in part no more than a paraphrase, in a slightly different terminology. It seems to me, therefore, that Jakobson (and Bolinger) must have completely misinterpreted the point of view expressed in [SS].

[21] In particular, Halle's *Sound Pattern of Russian* (The Hague, 1960). This is also the central topic in my "Transformational Basis for Syntax," to appear in the *Proceedings* of the IVth Texas conference, 1959, A. A. Hill, editor; and it is discussed in a section of [TAS].

[22] What led Hill to conclude that transformational grammar is concerned primarily with writing rather than speech was apparently my remark in [SS] that "in the context 'I saw a fragile—', the words 'whale' and 'of' may have equal (i.e., zero) frequency in the past linguistic experience of a speaker who will immediately recognize that one of these substitutions, but not the other, gives a grammatical sentence" (p. 16). The point in that context, was to show that judgment of well-formedness is not based on statistical order of approximation. This is rather obvious, and hardly deserves more than this casual comment. It is true that the phrasing of this observation makes use of an appeal to the literacy of the audience. It would have been possible to make the same point, more elaborately, without this appeal. That is, I could have stated that in the context Σ—(where Σ stands for a phonetic transcription of "I saw a fragile", with full intonation, etc.), /wéyl/ will be interpreted as the familiar lexical item *whale*, /əv/ will be rejected with stress 4, and, with stress 1, either rejected or treated as a new, unfamiliar lexical item, or as the quotation form *of*. I suppose that one could add even further qualifications, but there are other places where preciseness and rigor are likely to be more rewarding.

In [SS] examples (6) and (7) were used to illustrate the futility of (a) a search for a semantically based definition of "grammaticalness," or (b) an attempt to define grammaticalness in terms of order of approximation. Thus (6), but not (7), is clearly a well-formed sentence (or, at least, more like one than (7)), although (a) neither (6) nor (7) seems to be "meaningful" in any sense independent of grammatical structure and (b) the two are not differentiated in terms of statistical order of approximation. The same examples could be used to illustrate the impossibility of (c) defining "grammaticalness" in terms of a frame of "grammatical morphemes," since (8) has the same frame as (6) and (9) the same frame as (7) (note that it would be begging the question to argue that the morphemes *ly*, etc., are different in (6) and (8))

(8) harmless seem dogs young friendly (—less — —s — —ly)
(9) friendly young dogs seem harmless (—ly — —s — —less).

What Jakobson points out is that we can recognize the grammatical structure of (6) (but not (7)), and that "these grammatical relations create a meaningful sentence." He also observes that only (6) could appear in a poem,[23] that we know how to question it, that we can recognize in it certain metaphors, etc. These remarks are precisely in support of the argument in [SS] that (6) but not (7) is grammatical, and that the basis for whatever meaningfulness we can assign to it is its independently recognizable grammatical structure. The fact that the independently recognized "grammatical relations create a meaningful sentence" is perfectly consistent with the claim in [SS] that grammaticalness is not attributed to an utterance by virtue of the fact that the utterance is recognized to be meaningful. To argue against this claim, one would have to maintain that there is an absolute semantic property of "meaningfulness" that can be assigned to utterances quite independently of any consideration of their grammatical

[23] Hill, on the contrary, considers that (7) but not (6) could be a line of modern poetry ([G], p. 4). Both Hill and Jakobson refer to poems written to prove these (opposing) points. In fact, the question whether a sequence of words might appear in a poem is entirely beside the point, since it is perfectly plain that deviation from well-formedness is not only tolerable, in prose or poetry, but can even be effectively used as a literary device. We return to this below. It is by no means a novel observation. Cf., e.g., Empson, *Seven Types of Ambiguity* (Meridian), p. 34: "The demands of metre allow the poet to say something which is not normal colloquial English, so that the reader thinks of the various colloquial forms which are near to it, and puts them together; weighting their probabilities in proportion to their nearness. It is for such reasons as this that poetry can be more compact, while seeming to be less precise, than prose. It is for these reasons, too, among others, that an insensitivity in a poet to the contemporary style of speaking . . . is so disastrous . . ."

structure—a property that can be shown to hold of (6) but not (7). It is difficult to believe that anyone would seriously uphold this view.

From consideration of such examples as (6), (7) (and many others of different kinds—cf. [SS], §9) it seems evident that perception of grammatical relations in a sentence does not depend on a prior and independent identification of semantic properties, and that the study of grammatical structure seems to be, in fact, quite independent of meaning. Furthermore, as noted in [SS], p. 101,

> "it seems that the study of meaning is fraught with so many difficulties even after the linguistic meaning-bearing elements and their relations are specified that any attempt to study meaning independently of such specification is out of the question.[24] To put it differently, given the instrument language and its formal devices, we can and should investigate their semantic function (as, e.g., in R. Jakobson, "Beitrag zur allgemeinen Kasuslehre," *TCLP* (1936)); but we cannot, apparently, find semantic absolutes, known in advance of grammar, that can be used to determine the objects of grammar in any way."

Nothing in Jakobson's remarks gives any reason to revise these conclusions. He is saying that having recognized a sentence as grammatical and having identified its grammatical relations, a speaker can proceed to interpret it, determining how these formal elements and constructions are functioning, in this case. It will, in general, be easier to invent some sort of an interpretation for a grammatical than for an ungrammatical sequence, as Jakobson observes. This is basically the point of view expressed in [SS].[25]

In this connection, we can turn to Bolinger's discussion of what he regards as the semantic basis for transformational grammar ([LSLE], p. 377). He asserts that "knowing the meaning of *seem*, we can predict with some assurance that the structure **He is seeming* is not likely to occur." Under one interpretation, this remark is perfectly true. In just the same way we

[24] This is taken for granted in investigations that are actually concerned with meaning. Cf., e.g., Ziff, *Semantic Analysis* (Ithaca, 1960).

[25] The temporal implications in this description should not be taken too literally. However, I think that a reasonable perceptual model would have the property that grammatical structure is identified independently of any semantic consideration, and a reasonable model for language learning wou'd have the parallel property. This is a large topic in itself, and I will not attempt to pursue it here. I would merely like to emphasize that those who regard semantics as providing the basis, in some sense, for grammar, may not have realized how extreme and implausible is their claim, if stated quite openly. The claim must be that there are semantic absolutes, identifiable in noises independently of the grammatical structure (in particular, the lexical items) assigned to them. Perhaps some weaker claim than this is intended, but if so, it has never, to my knowledge, been clearly formulated.

could say that knowing the meaning of *seem*, we can predict that this element will not appear in the context *I saw—boy*, or knowing the meaning of *in*, we can predict the ungrammaticalness of *I saw in boy*, etc.[26] The comment is useful, however, only if we are presented with some characterization of the meaning of *seem*, *in*, etc., stated without reference to their grammatical function, from which this function can be predicted.[27] In the absence of this, Bolinger's statement reduces to the observation, misleadingly expressed, that *seem* does not appear in the progressive. In general, study of the continuing controversy over the reliance of grammar on meaning seems to me to show that the issue is only verbal—that no substantive claim is being made by those who claim that such reliance is essential.

In apparent conflict with [SS], Jakobson also argues that the notion "degree of grammaticalness" is untenable. Here too, however, I can find only a terminological issue. Jakobson recognizes ([BVMG], final paragraph) that there is some sort of scale of "obliteration" of "syntactic forms and the relational concepts which they carry," and he would, of course, insist on the importance of the distinction between literal and figurative usage. It is precisely such observations as these that motivate the study of deviation from grammaticalness to which he expresses his objection. Clearly the only question is whether the study of these problems is a natural outgrowth of grammatical investigations—that is, the question is to what extent considerations of the kind that are fundamental to the study of principles of sentence formation can provide some systematic account of such phenomena. I think that to some extent they can, and the investigation of degrees of grammaticalness (or, if one prefers, the scale of obliteration of syntactic forms . . .) is an attempt to show how a simple generalization of some familiar notions can give some insight into these problems.

5. Since the point has been widely misunderstood, I would like to emphasize that I am using the terms 'grammatical' and 'degree of grammaticalness' in a technical sense (which is, however, not unrelated to the ordinary one). In particular, when a sentence is referred to as

[26] The phrase "likely to occur" is often used in place of some such technical term as "grammatical." Its use gives an appearance of objectivity which is, however, quite misleading. In the literal sense of these words, practically nothing can be predicted about what is likely to occur in speech.

[27] It should be emphasized again that in studies devoted to meaning, the argument is usually reversed. Thus it is common to find the claim that knowing, seeing, etc., are not processes supported by the observation that *know*, *see* (like *seem*) do not naturally occur in the progressive, in their usual senses.

semi-grammatical or as deviating from some grammatical regularity, there is no implication that this sentence is being "censored" ([BVGM], p. 144) or ruled out, or that its use is being forbidden. Nor, so far as I can see, are there any "ontological" considerations involved ([LSLE], p. 377, [BVGM], p. 144), except insofar as these are reflected in grammatical categories and subcategories. Use of a sentence that is in some way semi-grammatical is no more to be censured than use of a transform that is remote from the kernel. In both cases, what we are attempting to do is to develop a more refined analysis of sentence structure that will be able to support more sophisticated study of the use and interpretation of utterances. There are circumstances in which the use of grammatically deviant sentences is very much in place. Consider, e.g., such phrases as Dylan Thomas' "a grief ago,"[28] or Veblen's ironic "perform leisure." In such cases, and innumerable others, a striking effect is achieved precisely by means of a departure from a grammatical regularity.

Given a grammatically deviant utterance, we attempt to impose an interpretation on it, exploiting whatever features of grammatical structure it preserves and whatever analogies we can construct with perfectly well-formed utterances. We do not, in this way, impose an interpretation on a perfectly grammatical utterance (it is precisely for this reason that a well-chosen deviant utterance may be richer and more effective—cf. footnote 23). Linguists, when presented with examples of semi-grammatical, deviant utterances, often respond by contriving possible interpretations in constructed contexts, concluding that the examples do not illustrate departure from grammatical regularities. This line of argument completely misses the point. It blurs an important distinction between a class of utterances that need no analogic or imposed interpretation, and others that can receive an interpretation by virtue of their relations to properly selected members of this class. Thus, e.g., when Jakobson observes ([BVGM], p. 144) that "golf plays John" can be a perfectly perspicuous utterance, he is quite correct. But when he concludes that it is therefore as fully in accord with the grammatical rules of English as "John plays golf," he is insisting on much too narrow an interpretation of the notion "grammatical rule"—an interpretation that makes it impossible to mark the fundamental distinction between the two phrases. The former is a perspicuous utterance precisely because of the series of steps that we must take in interpreting it—a series of steps that is initiated by the recognition that this phrase deviates from a certain grammatical rule of English, in this case, a selectional rule that determines the grammatical categories of the subject and

[28] One of the examples analyzed in Ziff's interesting study of the problem of deviation from grammaticalness, "On Understanding 'Understanding'" (mimeographed, 1960).

object of the verb "play." No such steps are necessary in the case of the nondeviant (and uninteresting) "John plays golf."

I am not, of course, suggesting that every difficult, interesting or figurative expression is semi-grammatical (or conversely). The important question, as always, is to what extent significant aspects of the use and understanding of utterances can be illuminated by refining and generalizing the notions of grammar. In the cases just mentioned, and many others, I think that they can. If this is true, it would be arbitrary and pointless to insist that the theory of grammatical structure be restricted to the study of such relatively superficial matters as agreement, inflectionally marked categories, and so on.[29]

In short, it seems to me no more justifiable to ignore the distinctions of subcategory that give the series "John plays golf," "golf plays John," "John plays and," than to ignore the rather similar distinctions between seeing a man in the flesh, in an abstract painting, and in an inkblot. The fact that we can impose an interpretation in the second case and sometimes even the third, using whatever cues are present, does not obliterate the distinction between these three strata.

Examples such as these provide a motive for the study of degrees of grammaticalness. Thus in addition to such data as (1), we can try to account for the observation that such phrases as (10) are not as extreme in their violation of grammatical rules as (11), though they do not conform to the rules of the language as strictly as (12):

(10) a grief ago; perform leisure; golf plays John; colorless green ideas sleep furiously; misery loves company; John frightens sincerity; what did you do to the book, understand it?

(11) a the ago; perform compel; golf plays aggressive; furiously sleep ideas green colorless; abundant loves company; John sincerity frightens; what did you do to the book, justice it?

(12) a year ago; perform the task; John plays golf; revolutionary new ideas appear infrequently; John loves company; sincerity frightens John; what did you do to the book, bite it?

[29] Notice that if we do, arbitrarily, limit the study of grammar in this way, we cannot even account for the difference between (6) and (9), on the one hand, and (7) and (8), on the other, since this difference can be expressed only in terms of categories that are established in terms of syntactic considerations that go well beyond inflection. But if we distinguish (6) from (9) by rules involving such syntactic categories as Adjective, Noun, etc., we can just as well distinguish "John plays golf" from "golf plays John" by rules involving such syntactic subcategories as Animate Noun, etc. These are simply a refinement of familiar categories. I do not see any fundamental difference between them. No general procedure has ever been offered for isolating such categories as Noun, Adjective, etc., that would not equally well apply to such subcategories as are necessary to make finer distinctions. I return to this below.

Here too, we can find innumerable relatively clear cases, and we can attempt to express these distinctions in a generative grammar (and, more importantly, we can try to find some basis for them through the study of generative grammar).

The question then arises: by what mechanism can a grammar assign to an arbitrary phone sequence a structural description that indicates its degree of grammaticalness, the degree of its deviation from grammatical regularities, and the manner of its deviation (cf. §1). This is a natural question to ask within the framework of §2.

Suppose that we have a grammar that generates an infinite set of utterances with structural descriptions. Let us call the units in terms of which these utterances are represented by the neutral term *formatives* (following a suggestion of Bolinger's). Suppose, in addition, that we have an m-level hierarchy of categories of formatives with the following structure. On level one we have a single category denoted C_1^1, the category of all formatives. On level two, we have categories labelled $C_1^2, \ldots, C_{n_2}^2$. On level three, we have categories $C_1^3, \ldots, C_{n_3}^3$, where $n_3 > n_2$, and so on, until we reach the m^{th} level with categories $C_1^m, \ldots, C_{n_m}^m$ $(1 < n_2 < \ldots < n_m)$. On each level, the categories are exhaustive in the sense that each formative belongs to at least one, perhaps more (in the case of grammatical homonymy). We might also require that each level be a refinement of the preceding one i.e., a classification into subcategories of the categories of the preceding level.

Let us assume, furthermore, that the m^{th} level categories are the smallest categories that appear in the rules of the generative grammar. That is, the members of C_1^m are mutually substitutable in the set of generated utterances. Many of them may contain just a single formative.

For concreteness, think of the formatives as English words.[30] Suppose we have a three-level hierarchy. Then C_1^1 is the class of all words. Let C_1^2 = Nouns, C_2^2 = Verbs, C_3^2 = Adjectives, C_4^2 = everything else. Let $C_1^3, \ldots, C_j^3$ be subcategories of Verbs (pure transitives, those with inanimate objects, etc.); subcategories of Nouns, and so on. Every sequence of words can now be represented by the sequence of first level, second level, third level categories to which these words belong. Thus "misery loves company" is represented $C_1^1C_1^1C_1^1$ on level one, $C_1^2C_2^2C_1^2$ (i.e., NVN) on level two, $N_{abstr}V_kN_{abstr}$ on level three (where these are the appropriate C_i^3's). One of the selectional rules of the generative grammars (i.e., in the transformational model of [SS], one of the context-restricted constituent structure rules) will specify that V_k occurs only with animate subjects. Thus "misery loves company" will not be generated by the

[30] This is merely an illustrative example.

grammar, though "John loves company" will. However, "misery loves company" has a level two representation in common with a generated utterance, namely, NVN. We therefore call it semi-grammatical, on level two. "Abundant loves company," on the other hand, has only a level one representation in common with a generated utterance, and is therefore labelled completely ungrammatical.

Without going into details, it is obvious how, in a similar way, a degree of grammaticalness can be assigned to any sequence of formatives when the generative grammar is supplemented by a hierarchy of categories. The degree of grammaticalness is a measure of the remoteness of an utterance from the generated set of perfectly well-formed sentences, and the common representing category sequence will indicate in what respects the utterance in question is deviant.[31] The more narrowly the m^{th} level categories circumscribe the generated language (i.e., the more detailed the specification of selectional restrictions) the more elaborate will be the stratification of utterances into degrees of grammaticalness. No utterances are "lost" as we refine a grammatical description by noting more detailed restrictions on occurrence in natural sentences. By adding a refinement to the hierarchy of categories, we simply subdivide the same utterances into more degrees of grammaticalness, thus increasing the power of the grammar to mark distinctions among utterances.[32]

Thus a generative grammar supplemented with a hierarchy of categories can assign a degree of grammaticalness to each sequence of formatives. If we could show how a hierarchy of categories can be derived from a generative grammar, then the latter alone would assign degree of grammaticalness. There are, in fact, several ways in which this might be possible.

Notice, first, that a transformational grammar will have such symbols as Noun, Adjective, . . . (in addition to much narrower subcategories) at

[31] We can represent only one "dimension" of deviation from grammaticalness in this way. There are others. Cf., e.g., [SS], §5, footnote 2. In obvious ways, we could give a more refined stratification of utterances by considering their parts, but I will not go into this.

[32] What is the natural point where continued refinement of the category hierarchy should come to an end? This is not obvious. As the grammatical rules become more detailed, we may find that grammar is converging with what has been called logical grammar. That is, we seem to be studying small overlapping categories of formatives, where each category can be characterized by what we can now (given the grammar) recognize as a semantic feature of some sort. If this turns out to be true in some interesting sense when the problem is studied more seriously, so much the better. This will show that the study of principles of sentence formation does lead to increasingly deeper insights into the use and understanding of utterances, as it is continually refined.

intermediate levels of representation, even if it is designed to generate only a narrow class of highly grammatical sentences, since these larger categories will simplify the descriptions of the domains of transformational rules. Thus we can expect to find a hierarchy of categories embedded within the constituent structure rules of the transformational grammar. This might be the appropriate hierarchy, or a step towards its construction.[33]

We might approach the question of projecting a hierarchy of categories from a set of utterances in a different way, by defining "optimal k-category analysis," for arbitrary k. Suppose, for simplicity, that we have a corpus of sentences all of the same length. Let $C_1, \ldots, C_k$ be (perhaps overlapping) categories that give an exhaustive classification of the formatives appearing in the corpus. Each sentence is now represented by at least one category sequence. Each such category sequence, in turn, is the representation of many sequences of formatives, in particular, of many that may not be in the original corpus. Thus a choice of k categories extends the corpus to a set of sentences that are not distinguishable, in terms of these categories, from sentences of the corpus. It is natural to define the optimal k-category analysis as that which extends the corpus the least, i.e., which best reflects substitutability relations within the corpus. Given, for each k, the optimal k-category analysis, we might select the optimal k-category analysis as a level of the hierarchy if it offers a considerable improvement over the optimal k-1-category analysis, but is not much worse than the optimal k+1-category analysis (this could be made precise, in various ways). It is easy to see that there are circumstances under which the optimal k-category analysis might contain overlapping classes (homonyms)[34] It is also easy to drop the restriction that all sentences be of the same length, and that the

[33] This possibility was suggested by some remarks of R. B. Lees.

[34] In general, it is to be expected that overlapping of categories will lead to an extension of the set of generated sentences, since categories will now be larger. Therefore, in general an analysis with disjoint categories will be preferred, by the evaluation procedure suggested above, over an analysis with an equal number of overlapping categories. Suppose, however, that the overlap includes true homonyms—suppose, e.g., that the categories N and V are allowed to overlap in such elements as /riyd/ (*read, reed*), etc. We now have two ways of representing the sentences *read the book* (namely, VTN or NTN), *the reed looks tall* (TNVA or TVVA), and so on, instead of just one (e.g., VTN and TVVA, if /riyd/ is assigned to V). We can select, in each case, the representation which is required, on independent grounds, by other sentences, i.e., VTN and TNVA, in this example. In this way we can reduce the number of generated sentences by allowing categories to overlap. Overlapping of categories will be permitted, then, when the gain that can be achieved in this way more than compensates for the loss resulting from the fact that categories are larger. We might inquire then whether homonyms can be defined as elements that are in the overlaps in the optimal set of categories on some level. Some evidence in favor of this assumption is presented in [LSLT].

corpus be finite. Such suggestions as these, when made precise,[35] offer an alternative way in which the generative grammar itself may impose degrees of grammaticalness on utterances that are not directly generated, through the intermediary of the category hierarchy projected from the set of generated sentences.

This suggestion is schematic and no doubt very much oversimplified. Nevertheless, such an approach as this to the problem of defining syntactic categories has many suggestive features, and offers some important advantages over the alternatives (e.g., substitution procedures)[36] that have been described in the literature (cf. [LSLT] for a detailed discussion—in particular, it allows for the possibility of setting up a hierarchy of categories and subcategories and for a principled and general solution to the problem of recognizing homonyms). I mention it here to indicate one way in which the further investigation of deviation from grammaticalness might be systematically pursued.

Massachusetts Institute of Technology
Cambridge 39, Mass.

[35] This approach to degrees of grammaticalness was described in more detail in [LSLT]. It was presented, with some supporting empirical evidence, in a Linguistic Institute lecture in Chicago in 1954, and again in the discussions of the IV[th] Texas conference, 1959.

[36] It is often proposed that categories be defined in terms of particular sets of inflectional morphemes, but unless some general method is given for selecting the relevant sets (none has ever been proposed, to my knowledge), such definitions are completely ad hoc, and simply avoid the problem of discovering the basis for categorization.

L. C. N. Y. Publications

MONOGRAPH SERIES

The *Linguistic Circle of New York* publishes occasional Monographs as supplements to various volumes of WORD. These Monographs are distributed gratis to members *at the time of publication.* They are available thereafter only on specific order and at the prices listed.

Monograph No. 1, Supplement to Volume 7, WORD
KARL HEINRICH MENGES—The Oriental Elements in the Vocabulary of the Oldest Russian Epos, The Igor Tale ..... *out of print*

Monograph No. 2, Supplement to Volume 9, WORD
YURY SERECH—Problems in the Formation of Belorussian **$2.50**

Monograph No. 3, Supplement to Volume 12, WORD
ANNA GRANVILLE HATCHER—Theme and Underlying Question; Two Studies of Spanish Word Order **$2.50**

SPECIAL PUBLICATIONS

1. URIEL WEINREICH—*Languages in Contact* (1953), xiv+146 pp. ..... *out of print*
2. *Linguistics Today* (1954), 280 pp. ..... *out of print*
   Equivalent to WORD, vol. 10, no. 2–3, but with hard-cover binding and separate pagination.
3. *The Field of Yiddish; Studies in Language, Folklore, and Literature*, ed. URIEL WEINREICH (1954), ix+317 pp. **$5.00**
   For members and subscribers (Please mark your order "Member Rate") **$4.00**

Complete sets of back numbers of WORD are no longer available. A list and quotation of those still in print will be sent on request. Information about reprinting will be found on the outside of the back cover.

Write to: JOHN P. HUGHES, St. Peter's College, 2641 Hudson Blvd., Jersey City 6, N.J.

ANNOUNCING

A SUBSCRIPTION CAMPAIGN

FOR REPRINTING BACK NUMBERS

The Linguistic Circle of New York is happy to announce that it has completed arrangements for reprinting the Monographs, Special Publications, and back numbers of WORD that are presently out of print, and virtually unobtainable.

Reprinting will be on a subscription basis. Those desiring back numbers or one of the Publications should send their orders and remittances to the Treasurer of The Linguistic Circle of New York (John P. Hughes, St. Peter's College, Jersey City 6, New Jersey). All orders will be acknowledged, and all remittances held in a special fund. When a sufficient sum is in hand for reprinting any number, it will be sent to press. If by December 31, 1962, sufficient orders have not been received to warrant reprinting any particular number, all remittances for it will be refunded as soon as possible after that date.

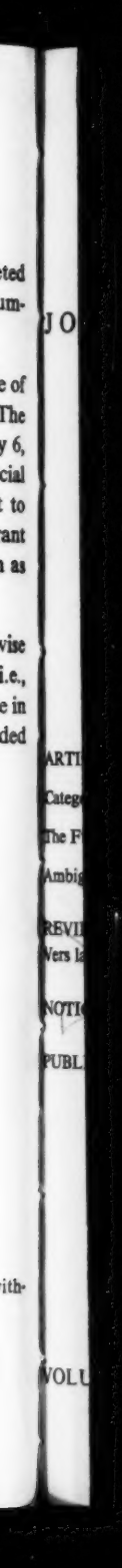
Following is a list of the publications presently out of print. Where not otherwise indicated, each single number is priced at $3.00, each double number $6.00, i.e., $9.00 for the ordinary volume of three numbers. (This includes postage anywhere in the world.) In cases where a bill is required, a pro-forma invoice will be forwarded on request.

WORD

Volume		*Number*	*Volume*		*Number*
II	(1946)	1–3	VIII	(1952)	1–4
III	(1947)	1–3		(No. 4=*Slavic Word* #1)	
IV	(1948)	1–3	IX	(1953)	2
V	(1949)	1–3	X	(1954)	1–4
VI	(1950)	1–3		(No. 4=*Slavic Word* #3)	
VII	(1951)	1–3	XII	(1956)	1–3
VII	*Supplement* (*Monograph* #1)		XIII	(1957)	1–2
				(All subsequent numbers in print).	

PUBLICATIONS

Linguistics Today (A separately paged and bound reprint of WORD X, No. 2–3) $5.00

NOTE: *Languages in Contact*, by Uriel Weinreich, formerly No. 1 in this series, has been withdrawn. A revised edition will be issued by another publisher in 1962.

PRINTED IN GREAT BRITAIN BY WILLIAM CLOWES AND SONS, LIMITED, LONDON AND BECCLES